ENGINEERING EUKARYOTIC-LIKE PROTEIN GLYCOSYLATION IN *ESCHERICHIA COLI*

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Juan Daniel Valderrama-Rincon

May 2012

ENGINEERING EUKARYOTIC-LIKE PROTEIN GLYCOSYLATION IN *ESCHERICHIA COLI*

Juan Daniel Valderrama-Rincon, Ph.D.

Cornell University 2012

N-linked glycosylation is a common protein post-translational modification where glycans are attached to asparagine residues located on a consensus sequence. Structure of these glycans varies widely among species, which is of particular importance given the role that glycans play on protein folding, functionality and recognition. Notably, glycans are often necessary to maintain stability and efficacy of therapeutic proteins in the human body.

Production of therapeutic glycoproteins in *Escherichia coli* has the potential to become a flexible and relatively cheap alternative to current production systems, even though *E. coli* has no native protein glycosylation machinery. Recently, the glycosylation machinery from *Campylobacter jejuni*, a gram negative bacterium, was transferred to *E. coli*, making protein glycosylation possible for the first time in this host. Unfortunately glycans synthesized by *C. jejuni*, and bacteria in general, are not appropriate for use in human therapeutics, hence engineering the synthesis of human-like glycans in *E. coli* becomes necessary.

This study is focused on recombinant expression and engineering of eukaryotic glycosyltransferases to enable *E. coli* to synthesize eukaryotic-like glycans suitable for in vivo production of therapeutic glycoproteins.

Initially, we expressed glycosyltransferases from *Saccharomyces cerevisiae* (Alg13, Alg14, Alg1 and Alg2) for the synthesis of a core glycan structure (trimannose core) common to many human glycans. Later, the work was ori-

ented towards the extension of this trimannose core for production of authentic human-like glycans and improvement of the current system using directed evolution of glycosyltransferases. In the future, we expect to be able to produce glycosylated proteins in vivo, displaying fully sialylated human-like glycans.

BIOGRAPHICAL SKETCH

Juan Daniel Valderrama-Rincon was the first of three children born to José Joaquín Valderrama and María del Carmen Rincon. As the son of a mathematician and a biologist, his early interest on these fields led him to pursue a Chemical Engineering degree at Universidad Nacional de Colombia where he had the opportunity to start his research career on fuel biodesulfurization under the guidance of Dr. Edelberto Silva and Dr. Luis Caicedo, obtaining the national price for the best undergraduated thesis work in 2002 alongside with his partner Juan Carlos Cruz-Jiménez. Immediately after, he was accepted for a Master's Degree in Environmental Engineering at the same institution, where he worked with Dr. Maria Consuelo Díaz-Báez on anaerobic digestion of recalcitrant halogenated aromatic compounds. Later, he worked as lecturer and research associate at this same institution after which he was hired as director of the Innovation and Technological Development Program at Universidad Antonio Nariño. Two years later he was accepted for a Ph.D. at Cornell University in the Chemical and Biomolecular Engineering department where he joined Dr. Matthew DeLisa's lab and received his Doctorate of Philosophy in 2012.

Este trabajo está dedicado especialmente a mi esposa Andrea, quién me ha acompañado incondicionalmente durante estos cinco años y me ha brindado toda su paciencia y comprensión.

También agradezco a mis padres, Carmencita y Joaco, a mis hermanos, Carolina y Andrés, y en general a toda mi familia por tantos años de buenos momentos.

Finalmente, dedico este trabajo a la memoria de mi gran amigo Efrén Serrano, lamentablemente fallecido como victima de la guerra civil en mi país.

ACKNOWLEDGEMENTS

This work was supported by the National Science Foundation Career Award CBET-0449080 (to Dr. Matthew P. DeLisa), the New York State Office of Science, Technology and Academic Research Distinguished Faculty Award (to Dr. Matthew P. DeLisa), the National Institutes of Health Small Business Innovation Research grants R43 GM087766 and R43 GM086965 (to Dr. Adam C. Fisher), the National Institutes of Health National Center for Research Resources grant 1 P41 RR018502-01 (to the Complex Carbohydrate Research Center) and a graduate fellowship from LASPAU and the Universidad Antonio Nariño (to Dr. Juan D. Valderrama-Rincon).

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

The glycome, defined as the totality of carbohydrates in an organism, is considered as "one of the most complex entities in nature" [171] even surpassing the complexity of the genome and the proteome. The role of this glycome in relation to the living organisms is studied by glycobiology, a relatively recent field of biology that has been gaining importance due to its relation to other major fields such as immunology, biochemistry and bioengineering. The term "glycobiology" was used for the first time by biochemist Raymond Dwek at the University of Oxford, UK in 1988 [20]. The purpose of this term was to emphasize the role of carbohydrates and glycans in biological systems, as opposed to the study and analysis of these molecules as isolated entities unrelated to their biological activity [20]. The late arrival of glycobiology may be mainly attributable to the fact that the actual importance of glycans in processes like cell signaling, cell differentiation, immunological recognition and protein stability, has only been recognized recently [131]. Moreover, it was not uncommon to think about glycans and polysaccharides as just agents decorating cell surfaces and protecting them from proteolysis [20].

In addition to the glycome complexity, carbohydrate analysis and glycan analysis in general have always been challenging tasks [3, 24, 95, 131]. Among several reasons, the high variety of glycans, their complicated structures, and the similarity in their subunit components (monosaccharides) are the features more often cited as responsible for this difficulty [3, 131]. It has been even stated that technical difficulties related to glycobiology are often so high that it could be the reason why it is difficult for some scientists to get into the field [20]. Despite these difficulties that can be associated to any emerging field in science,

glycobiology has been constantly gaining momentum. We are right now at a point that our collective knowledge, though still very limited, has already permitted us to engineer biological systems around the synthesis of *ad-hoc* glycans. For example, cell surface display of glycans and protein glycosylation have been customized, taking advantage of current techniques in protein and metabolic engineering [21, 51, 60, 64, 78, 90, 141].

Based on this advances and given their importance in nature, in this work we focused mainly on glycans and its engineering on *E. coli* cells. Our goal was to achieve *in vivo* synthesis of eukaryotic-like and human-like glycans in this bacterial cells, keeping in main as a long term goal the use these engineered glycans as precursors for *in vivo* synthesis of genuine human glycoproteins in *E. coli*.

1.1 Glycans

Glycans are complex molecules made of monosaccharides. As shown in figure 1.1a, monosaccharides have several hydroxide groups available for reaction and polymerization. Because of this feature, glycans can be as simple as disaccahrides or as complex as highly branched molecules comprising more than 10 single units (figure 1.1b).

Given these characteristics, glycans constitute the largest class of biological macromolecules [169]. Indeed, it has been calculated that the potential variability of glycans surpasses that of nucleic acids and peptides by several orders of magnitude [93]. Interestingly, while the number of common monosaccharides (approximately 12) is less than the number of common amino acids (20), the fact that glycans can happen as branched molecules and each subunit have 3 or more functional groups available for polymerization, accounts for the higher

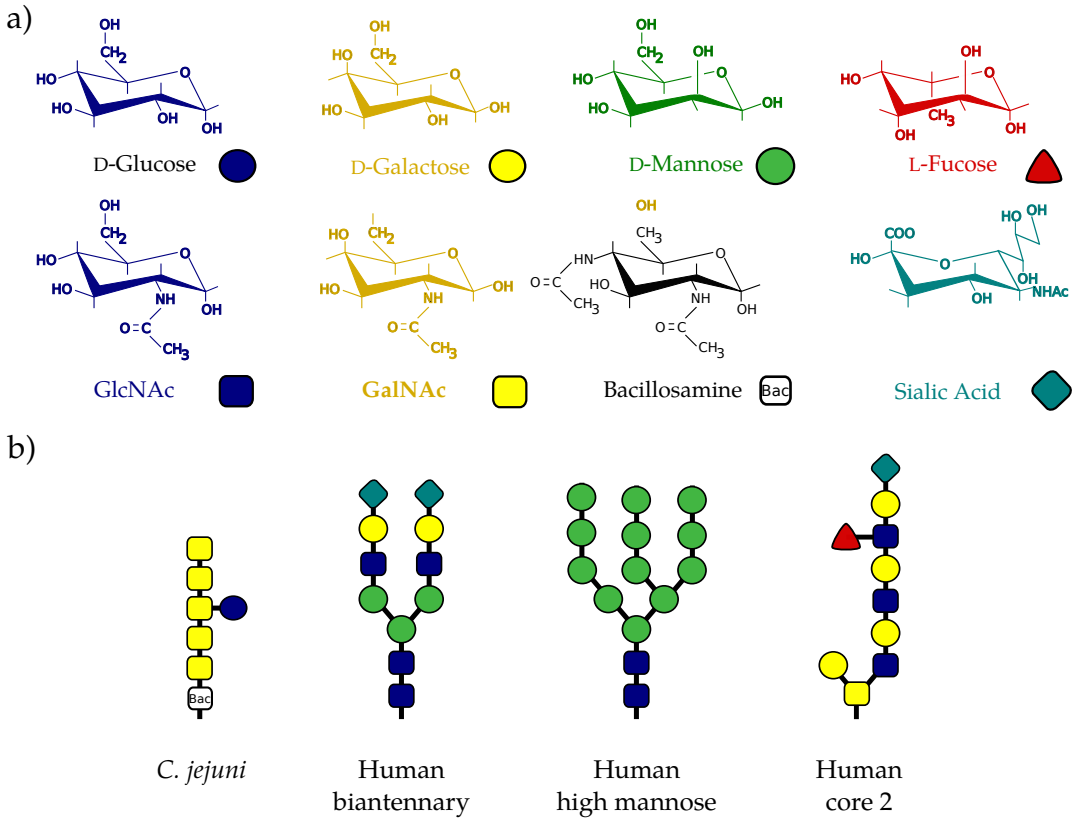

Figure 1.1: *Glycans in nature.* (a) Some monosaccharides commonly found as subunits in glycans. (b) examples of glycan variety in nature [108].

variability when compared with peptides and nucleic acids, which are restricted to a linear configuration (figure 1.2).

Thanks to this variability, glycans can store massive amounts of information. However, because of this variability, analysis and identification of glycans is still difficult mainly because monosaccharides are very similar among them [131]. For instance, glucose, mannose and galactose are just positional isomers where the only difference is the relative position of their hydroxide groups. Because of this factor, development of enhanced analytical methods applicable to carbohydrates is advancing constantly. Unfortunately many of the new techniques are still in their infancy and a fully standardized methodology for analysis is not available yet.

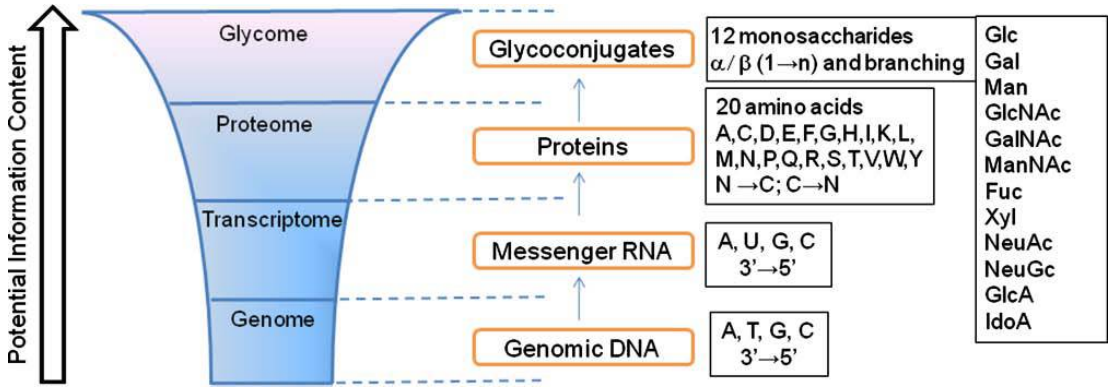

Figure 1.2: *Glycan variability*. Schematics comparing the potential size of the glycome in comparison with the proteome, the transcriptome and the genome (the diagram is not drawn to scale). Adapted from [62].

1.1.1 Glycan analysis

With glycobiology gaining more attention, analysis methods aimed at carbohydrates have been going through notorious improvement. Some traditional techniques, such as radiolabeling, are slowly being displaced by more convenient approaches like HPLC and capillary chromatography, while other classic techniques like fluorophore assisted carbohydrate electrophoresis (FACE), are being modified for enhanced results. Since any glycobiology related project is completely dependent on successful application of these techniques, a brief review of most prevalent methods for carbohydrate analysis is presented as follows.

Radiolabeling

Radiolabeling of glycan precursors is a method that has been applied to lipid linked oligosaccharide (LLO) analysis since nearly 30 years ago [140, 156]. In general, [2-^{3}H]mannose is fed directly to cells which can metabolize it into [2-^{3}H]mannose-6-P, followed by [2-^{3}H]mannose-1-P, and finally to GDP-[2-^{3}H]mannose [97]. This last compound, GDP-mannose or its radiolabeled version GDP-[2-^{3}H]mannose, is the activated donor for several glycosyltrans-

ferases; hence [2-^{3}H]mannose can be incorporated into a variety of glycans and exploited for further analysis [140].

Alternatively, other radiolabeled carbohydrates like [1-^{3}H]*N*-acetylglucosamine and [1-^{3}H]galactose [97, 138] can be used to target more specific glycans or as additional controls. In particular, [2-^{3}H]mannose is preferred for eukaryotic LLOs analysis, given the presence of multiple mannose subunits in most of those molecules, which results on stronger signals and hence higher sensitivity [97].

In general, the labeled sugars are added directly to cell culture during approximately 5min. Radioactive medium is then removed and lipids are extracted using chloroform and methanol. Thin layer chromatography (TLC) is commonly used for analysis of radioactive LLOs. Alternatively, glycans can be released from LLOs an then analyzed using liquid chromatography or FACE [140]. The main drawback of this approach resides precisely in the final analysis step; where methods like chromatography or electrophoresis can fail when trying to separate molecules that are very similar. Also, since this method can only provide limited information about glycan structure and composition, complementary analyses are required.

Fluorophore assisted carbohydrate electrophoresis (FACE)

Electrophoresis of glycans depend mainly on two characteristics: the electrophoretic mobility of glycans and the capacity of detecting and imaging these glycans directly in a gel. Many monosaccharides and glycans are not charged (they will not migrate during electrophoresis) or are just lightly charged [125], which together with their lack of chromophoric or fluorophoric groups, make them relatively hard to detect and not suitable for electrophoresis analysis [158].

These inconveniences have been addressed by conjugation of charged fluorophores which can tackle both the electrophoretic mobility and the detection problem, enabling glycans and carbohydrates for electrophoresis analysis [158]. In general, aromatic primary amines have been the fluorophores of choice, being gradually replaced by polysulfonic acid derivatives such as 8-aminonaphthalene-1,3,6-trisulfonate (ANTS) and 7-amino-1,3-naphthalenedisulfonic acid (ANDS) [56].

FACE analysis presents important advantages including the low cost of required reagents and equipment, relatively simple experimental set-up, and the fact that radioactive precursors are not required [55]. On the other hand, except for the addition of radioactive carbohydrates, the experimental setup for FACE is completely analogous to the setup previously described for radiolabeling, where cells are grown for a determined period of time and LLOs are then extracted for analysis. The main difference is found after glycans are released from LLOs by acid hydrolysis. When applying FACE, glycans are derivatized using the fluorophore of choice, by reductive amination, and then samples are loaded into a polyacrylamide gel for electrophoresis. Glycans can then be detected by fluorescence using a UV transilluminator [56, 97, 158]. Unfortunately, as is the case with radiolabeling, information about glycan structure and composition obtained using FACE is relatively limited and complementary approaches are necessary depending on the case.

Metabolic labeling of glycans with bioorthogonal modified carbohydrates

Recently, some modified carbohydrates have emerged as bioorthogonal compounds able to get incorporated into glycan structures without interfering with their metabolic activity and facilitating glycan detection and analysis by mean

of reactions like azide-alkyne (click chemistry) or azide-phosphine (Staudinger reaction). This method dates from 1992, when Reutter and coworkers modified *N*-acetylmannosamine replacing the acetyl group by an acyl chain homologue. After incorporation and metabolization of the modified carbohydrate, the acyl chain permitted detection and analysis of modified *N*-acetyl neuraminic acid, directly derived from modified *N*-acetylmannosamine inside cells [83, 94]. Further modifications to this technique include the use of ketones, thiol and azide groups in replacement of the acyl group, making the technique more flexible and extending its potential applications [94].

In general, azide sugars have been preferred over other modified carbohydrates based on azides high intrinsic reactivity, selectivity, stability in water, and unreactivity respect to most biological compounds [17]. Also, the azide moiety is small, making it ideal for carbohydrate modification [17, 95]. Its detection relies mainly on two approaches: click chemistry and the Staudinger reaction. Azide-alkyne cycloaddition is a typical click chemistry reaction, it is copper-catalyzed and has a rapid rate of reaction. However, its major drawback is copper toxicity and hence limited application on *in vivo* labeling [17]. On the other hand, the main application for Staudinger ligation is precisely in *in vivo* labeling. This reaction is particularly useful when a phosphine is attached to the azide group, being the advantage based on the fact that phosphine can be conjugated to a variety of tags such as biotin, FLAG, myc and His_6 facilitating considerably its detection [3, 94]. Besides that, the main limitation of these methods is related to the availability of modified sugars. Azide sugars can only be obtained from *N*-acetylate carbohydrates like *N*-acetylglucosamine (GlcNAc) or *N*-acetylgalactosamine (GalNAc) [95]. Because of this limitation, the method is restricted only to certain glycan structures.

Successful applications of bioorthogonal modified carbohydrates include labeling of fucosylated glycans in mammals, protein glycosylation studies in bacteria, tagging of *O*-GlcNAc modified proteins, and glycan analysis in yeast [22, 39, 86, 157, 190].

Lectin analysis and microarrays

Antibodies have been extensively used for protein detection and cell labeling mainly thanks to their relatively good availability, high specificity, and ease of use. In general, new antibodies can be obtained by inoculating a suitable mammal with the protein (antigen) of interest, then the animal's immune system may generate "new" antibodies against this protein and then these antibodies (IgGs in particular) can be further purified from the animal's serum for further use. Unfortunately in the case of glycans this protocol is no as useful as it is with proteins. The main challenge in this case is related to the fact that, in general, the immunogenicity of carbohydrates is relatively low [66], so antibodies against glycans are difficult to obtain using traditional methods.

Even though there are available antibodies against some glycans, lectins are more commonly used for glycan affinity-based detection. However, it is important to keep in mind that, compared to antibodies, lectins usually have relatively low affinity for their carbohydrate epitope and many times rely on multivalency for appropriate binding [95]. Despite these disadvantages, lectins have proven to be widely useful and have some of the advantages attributed to antibodies. For example, lectins can be used for cell surface labeling *in vivo*, detection of glycoproteins in an immunoblotting analogous way, labeling of tissues and organs, and agglutination analysis of bacterial and mammalian cells [49, 62, 68, 95, 152, 161].

In the context of this project, concanavalin A (ConA), a lectin extracted from *Canavalia ensiformis*, is of particular interest given its affinity towards mannose and several oligosaccharides displaying terminal mannose residues [114, 117, 136]. Because of this feature, it has been extensively used in tissue analysis. However, its importance for this project resides on its high affinity towards the "trimannose core", an eukaryotic like glycan which synthesis was considered as our main goal [114, 117]. This feature permitted us multiple kinds of analysis, like surface labeling of *E. coli* cells and detection of glycoproteins and glycolipids.

Recently, lectin analysis has evolved from the use of single lectins to the arrangement of several lectins into microarrays for high throughput analysis. Lectin microarray technology depends on immobilization of these proteins over a solid substrate. Immobilization methods tested so far include: carbene insertion, interaction between biotin and avidin, and use of amino acid side chains like lysine [62]. Applications of this technology are quickly increasing and include analysis of bacterial glycome and its response to metabolic alterations, glycan profiling and identification, glycan-based identification of bacterial pathogens, and clinical diagnostic [42, 71, 82, 91, 164].

Finally, it is worth to mention that lectin analysis provides some insight into glycan structure, given the specificity of lectins towards certain carbohydrate motives. Based on this factor, glycan microarrays have been applied alongside with mass spectrometry (MS) for the analysis of protein glycosylation profiles and glycan identification [62]. Still, the information obtained from this approach is often not enough for a complete identification and more sophisticated techniques are necessary.

Mass spectrometry (MS)

MS comprises a good number of different methods and techniques where a single one is probably not good enough for analyzing every single kind of glycan. Matrix-assisted laser desorption/ionization (MALDI) is the most widely used, given most glycans give signals under the conditions for this method even on their native states [65].

Often, glycans occur in nature conjugated to lipids and proteins. Because of that, when possible, it is preferred to release them before analysis to facilitate the interpretation of results [115]. When glycans are attached to proteins, enzymatic releasing is preferred, being PNGase F the most popular enzyme for this purpose [165]. On the other hand, releasing of glycans from lipids is usually achieved by acid hydrolysis followed by liquid-liquid extraction and drying [56, 80]. It is also important to take into account that glycans are not ionized as efficiently as other molecules like proteins, a reason why derivatization is frequently used before MALDI to obtain improved results. Permethylation with a methyl iodide is the most common approach for increasing ionization efficiency and hence enhanced detection [65]. Permethylation is also particularly useful for stabilization of sialylated glycans, given sialic acid is easily removed from these molecules by hydrolysis, making its analysis more difficult [115].

Recent advances on MS/MS analysis and MS^n techniques have provided further insight into glycan structure, having now the possibility of obtaining basic information respect to sequence, linkage, and branching based primarily on MS analysis [134, 150]. Nevertheless, despite all the advances on MS, most detailed analysis of glycans is still challenging. The main problem is still related to the similarity among monosaccharide subunits. In particular, it is not possible to discriminate structural isomers using MS, being only possible to characterize

these monosaccharides generically as pentoses or hexoses [104], for exmaple, so definitive identification still relies on NMR analysis.

Nuclear magnetic resonance (NMR)

As mentioned before, carbohydrate structural isomers cannot be identified using solely MS, being NMR analysis crucial for detailed glycan identification and analysis [46]. NMR is then the ultimate tool for glycan analysis, permitting a complete identification of both glycan structure and monosaccharide nature. Generally, NMR spectra is analyzed by expert personal, supported on specialized software and glycan structure databases [96].

Usually, glycan NMR analysis requires from a few micrograms to a couple miligrams of sample [23, 61], which, depending on the glycan source, could represent an additional challenge given sometimes is difficult to obtain enough amounts of sample. Still, it is generally considered that NMR is to certain point independent of the sample size [46].

Finally, it is worth to mention that publishing good NMR data has been gaining importance on the glycobiology field. Because of that, reporting tables of chemical shifts is a widespread practice and recommended when presenting NMR results [46].

1.2 Protein glycosylation

Glycosylation of proteins is the most common post-translational modification in nature. It is present in all three domains of life playing a role on protein localization, recognition, folding and quality control [107]. *N*-linked and *O*-linked glycosylation are the prevalent types of protein glycosylation, being *N*-linked

glycosylation the most abundant and the best understood. *N*-linked glycosylation refers to the covalent conjugation of a glycan molecule to an asparagine (N) residue in a protein. To be glycosylated, this asparagine residue has to be part of a recognition sequence of amino acids: N-X-S/T (where X is any amino acid but proline) in the case of eukaryotes [107]. In prokaryotes, the N-X-S/T sequon is not necessarily sufficient, like is the case for *Campylobacter jejuni* where the recognition sequence is extended to D/E-Z-N-X-S/T (where Z is any amino acid but proline) [88], being D-Q-N-A-T the optimal acceptor sequence [29] for this organism.

Interestingly, the *N*-linked glycosylation process shows homologous features in all three domains of life, for example, glycans are usually assembled stepwise by several glycosyltransferases over a membrane embedded lipid carrier. The resulting structure, a lipid-linked oligosaccharide (LLO) is then flipped to the luminal space of the endoplasmic reticulum (ER) in eukaryotes, or to the periplasmic space in prokaryotes. In the eukaryotic case, LLOs are further modified by incorporation of additional monosaccharides to the glycan structure. Finally, an oligosaccharyltransferase (OST) catalyzes *en bloc* transfer of the glycan to the recognition sequence in an acceptor protein [146]. Figure 1.3 shows a schematic of the glycosylation process in *C. jejuni*, which is in some way a simplified version of the eukaryotic glycosylation machinery.

The importance of protein glycosylation in nature is already well established. However, the exact function of most glycans and their actual impact on glycoproteins is still being investigated. From the point of view of this project, the main interest resides on the effect that glycans have on proteins with potential therapeutic applications and the possibility of producing these glycoproteins in *E. coli*.

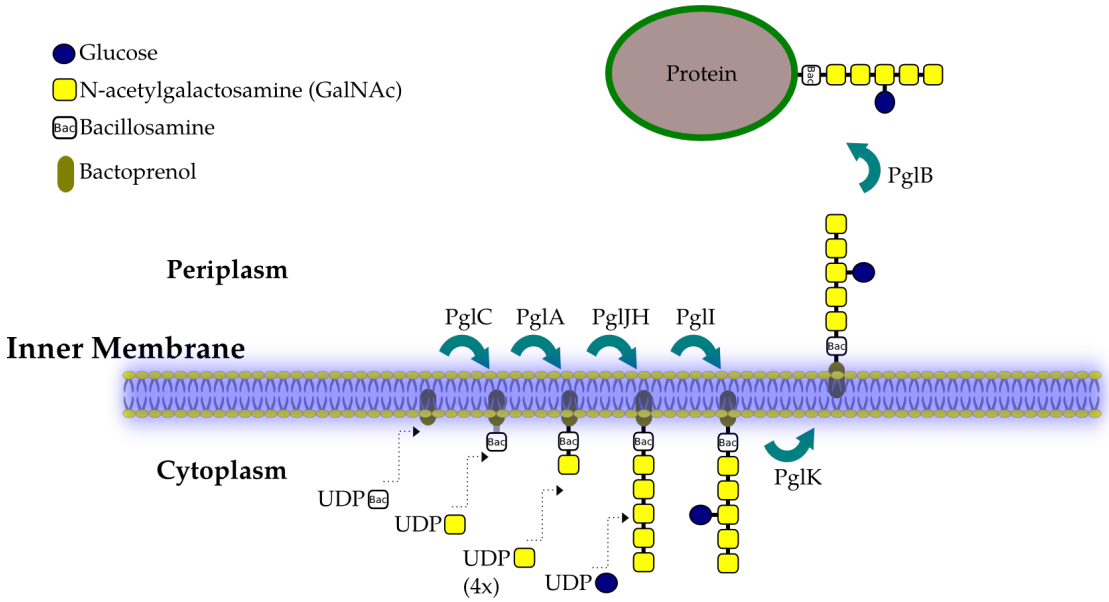

Figure 1.3: *Schematic of C. jejuni protein glycosylation pathway*. Adapted from [84].

1.2.1 Effects of glycosylation on therapeutic proteins

It has been recognized that proteins can display higher specificity and potency than small molecule based drugs. Unfortunately, their instability presents a major challenge for pharmaceutical applications [6], where the main issue is related protein structure (secondary and tertiary) mainly because thermodynamic stability of the protein active conformation is often not very different from the thermodynamic stability of some inactive conformations and hence proteins may easily lose activity under certain circumstances. Because of this problem, most of the current efforts are focused on strategies for maintaining an stable active conformation of these therapeutic proteins under physiological conditions [7].

Among several strategies available to protect these active conformations the most commonly used are the insertion of mutations in the amino acid sequence, use of excipients in the final formulation, protein pegylation, and protein glyco-

sylation. Particularly, glycosylation is desired, given it has been shown to have effects not just on stability and protein structure but also on the therapeutic efficacy of these proteins. It has also being suggested that all the beneficial effects of glycosylation are dependent on several characteristics like glycosylation degree (site occupancy), glycan size and glycan structure [154].

Glycans can also prevent protein aggregation mainly because of their hydrophilic nature, increasing the solvent accessible area of the protein and hence making glycoproteins more soluble [153]. Furthermore, it has also been found that glycosylation confers additional resistance to extreme pH, chemical denaturation, and thermal denaturation [154]. On the other hand, the *in vivo* stability of therapeutic proteins depends strongly on its resistance to proteolytic degradation [163] which has been demonstrated to be related to proper glycosylation [151]. It was suggested that this resistance is conferred due to the steric interference provided by the glycan, blocking direct contact of the protease with the protein backbone and preventing cleavage [15, 154]. It is also important to mention that glycans, and in particular terminal-sialylated glycans, affect serum half-life of proteins. One reason for this is that non-sialylated glycans with terminal galactose (Gal) or GlcNAc residues are recognized by the asialoglycoprotein receptor on the surface of hepatocytes, leading to protein internalization and degradation [15]. This fact makes evident that the protein glycosylation problem is not just limited to glycan conjugation, but instead, it is also related to glycan structure. For example, human cells produce only Neu5Ac-type sialic acid while other mammals produce the slightly different Neu5Gc-type, because of that, some glycoproteins produced in non-human cells happen to be immunogenic [126]. In a similar way, xenotransplantation of organs from pigs is prevented because Gal residues in pig glycans are attached by an α1,3 linkage

which also results immunogenic to humans [135].

Using bacterial glycans does not seem to be adequate either. In principle, *C. jejuni* glycosylation machinery could be exploited for therapeutic glycoprotein production in bacteria. However, the problem with this approach arises from evidence suggesting that the immunogenicity of certain glycoproteins from this bacterium is strongly related to the glycan moiety attached to them, making *C. jejuni* glycan unsuitable for use in humans [162].

In summary, therapeutic application of glycoproteins is gaining importance thanks to the accumulated evidence showing how glycans impact both stability and activity of proteins in humans. Nevertheless, it is not just a matter of attaching glycans to proteins, on the contrary, glycan localization and glycan structure are also important factors to take into account, adding to the complexity of the glycoprotein engineering problem.

1.2.2 *N*-linked glycans in eukaryotes

The protein glycosylation process has been, to some extent, conserved across all three domains of live. As mentioned before, the process is centered on LLO synthesis occurring at the membrane level and transferring of the whole glycan to a receptor protein by an OST, where the main differences between the eukaryotic and prokarytic processes are the glycan structures and the nature of the OST.

In regard to glycan structure, the investigations so far suggest that there is considerable variety on bacterial *N*-linked glycans and it is suspected that this variability is even higher for archaeal organisms. In contrast, the vast majority of eukaryotes synthesize the same glycan precursor: Glc_3-Man_9-$GlcNAc_2$ (glucose: Glc, mannose: Man, *N*-acetylglucosamine: GlcNAc), so the actual vari-

ability in eukaryotic glycans is a result of further modifications happening in the late ER or the Golgi [146]. Figure 1.4 shows a sample of such variety in human *N*-linked glycans.

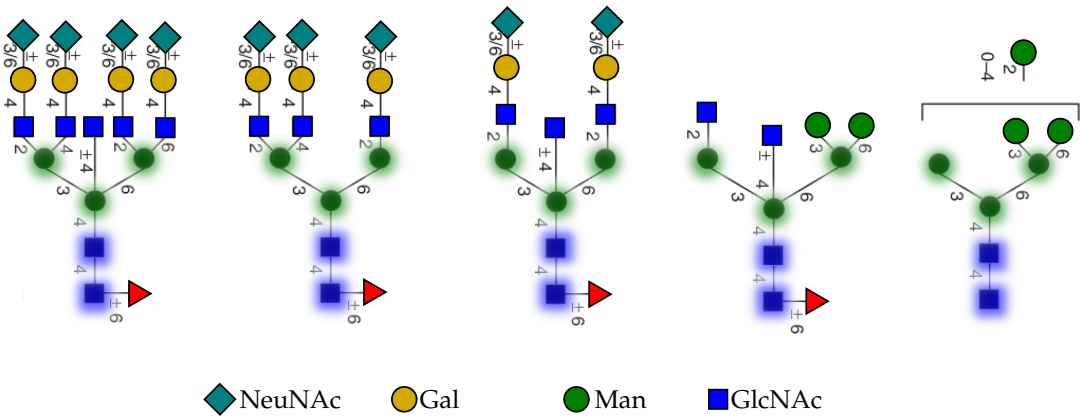

Figure 1.4: *Common glycans found in human glycoproteins*. The highlighted residues constitute the "common core" or "trimannose core" (adapted from [115]).

It is important to address how all structures in Figure 1.4 share a common "trimannose core". Notably, the trimannose core, as part of the Glc_3-Man_9-$GlcNAc_2$ precursor, is a conserved structure in eukaryotes which means also that its biosynthetic pathway is virtually the same from protozoans to humans [2, 146].

1.2.3 Engineering human-like *N*-linked protein glycosylation

By 2006 the therapeutic protein market was calculated on US $57 billion, constituting the the most important class of new products in the biopharmaceutical industry. Given proper glycosylation is indispensable in this field, mammalian cell expression systems are preferred because they are better suited for synthesis of human-compatible glycans. Among these expression systems can be listed Chinese Hamster Ovary (CHO) cells, mouse myeloma cells, human fibrosarcoma cells, human lymphoma, and human embryo kidney cells [45]. Never-

theless, these platforms present several drawbacks, like the need for expensive media, risk of viral contamination and relatively difficult genetic manipulation [38]. Because of that alternative systems are being investigated, being examples of these emerging systems, modified plant cells, filamentous fungi, insect cells and yeast cells [45].

In general, non human cells require some extent of engineering before proper glycosylation of proteins can be achieved. For instance, CHO cells can produce glycans resembling their human counterparts to a good extent. However, incomplete synthesis of humanized glycans in these cell lines is not rare, which implies that an important fraction of recombinant glycoproteins will not display terminal sialic acid. This issue has been addressed mainly by overexpression of β1,4-galactosyltransferase and α2,3-sialyltransferase to enforce synthesis of fully sialylated glycans but the results may depend on several other factors and are not always satisfactory [180].

In the case of plant cells, removal of β1,2-xylosylation and core α1,3-fucosylation is generally required. This engineered cells can synthesize human-like glycans lacking both terminal sialic acid and galactose, so additional engineering is still required for production of completely humanized glycoproteins in this system [26].

The insect cell-baculovirus system is also promising for glycoprotein production. Fully sialylated glycans can be synthesized by recombinant expression of β1,4-galactosyltransferase and α2,6-sialyltransferase. The main problem here reside in the fact that insect cells synthesize mostly monoantennary glycans, an issue that can be solved by recombinant expression of *N*-acetylglusaminyltransferase II. Additional challenges are the competitive production of highly mannosylated and terminally paucimannosylated glycans

[69].

One of the most recent advances in production of human-like glycoproteins is based on glycoengineering of *Pichia pastoris*. The first requirement in this case is eliminating the synthesis of highly mannosylated (and highly antigenic) glycans by deletion of the *OCH1* gene, which produces a Man_8-$GlcNAc_2$ structure. Recombinant expression of mannosidases permit the generation of the trimannose core that can be further elongated by recombinant expression of β1,2-*N*-acetylglucosaminyltransferase I and II, β1,4-galactosyltransferase and α2,6-sialyltransferase. Nevertheless, some issues still to be solved are related to undesired *O*-linked glycosylation and heterogeneity in glycan structure [78].

Compared to the already mentioned systems, engineering of a bacterial based system is still on its early steps. This is mainly due to the lack of protein glycosylation machinery in the majority of bacteria, limiting the availability of native pathways for further engineering. The most successful approach to date relies on the fully functional transfer of the protein glycosylation machinery from *C. jejuni* to *E. coli* [173]. Based on this advance, recombinant expression of a number of bacterial glycosyltransferases had permitted the synthesis of a diversity of glycans that can be successfully attached to proteins based on *C. jejuni* PglB enzymatic activity [51]. Despite this advances, the bacterial system still lacks the ability to produce more appropriated human-like glycans, being among the most important reasons the fact that bacterial glycans are substantially different from their eukaryotic counterparts [146], hence engineering the synthesis of human-like glycans in bacteria can not be based on existing bacterial pathways but instead requires the assembly of a new synthetic pathway.

The main goal of this project consisted on engineering the synthesis of the trimannose core in bacteria, serving as proof of principle and starting point for

engineering the synthesis of fully sialylated human-like glycans in *E. coli*, contributing to the solution of the glycan problem in bacterial systems for production of therapeutic proteins.

1.3 The trimannose core

The virtual ubiquity of the trimannose core in eukaryotic *N*-glycans implies that the synthesis of most human *N*-linked glycans can be achieved by extension of this relatively simple structure. Taking this into account, engineering the synthesis of the trimannose core in *E. coli* constitutes both, a proof of concept for the technical possibility of synthesizing more complex glycans, and an obligated first set of reactions preceding the synthesis of fully humanized glycans.

To date, it has not been reported a set of bacterial glycosyltransferases that would permit the synthesis of the trimannose core. This limits the possibilities to the eukaryotic domain where these glycosyltransferases have been classified as part of the Asparagine linked glycosylation (Alg) pathway and consequently all of them share the collective "Alg" designation [25]. Given the best studied members of these Alg group of proteins are from *Saccharomyces cerevisiae*, glycosyltransferases from this organism are the obvious first candidates for the engineering of the trimannose core synthesis pathway in bacteria. Additional reasons include the availability of the genetic material, the lack of introns in the specific genes encoding for these glycosyltransferases which facilitates the cloning process, and preceding studies showing that some of these Alg enzymes from *S. cerevisiae* could be expressed in *E. coli* [35, 122]. The following subsections provide a brief description of each of the *S. cerevisiae* glycosyltransferases involved in the synthesis of the trimannose core, as well as additional informa-

tion respect to the expression of these enzymes in *E. coli*.

1.3.1 Alg7

The first step in the synthesis of eukaryotic glycans is the attaching of of phospho-GlcNAc to a lipid carrier (dolichol) which is embedded in the membrane of the ER [25]. This reaction is catalyzed by the GlcNAc-1-phosphate transferase Alg7 [133]. Probably Alg7 is the less studied glycosyltransferase in the trimannose core synthesis pathway. There is no experimental data about its membrane topology, although hydrophobicity analyses predict at least 8 transmembrane domains [142]. Interestingly, Alg7 shows relatively high positional identity with its bacterial and archaeal homologs, suggesting the presence of Alg7 homologs in all three domains of life [142].

This homology is specially important when taking into account Alg7 theoretical membrane topology. It has been reported extensively the difficulty in recombinant expression of membrane associated proteins, particularly when those proteins are from eukaryotic origin and the target host is a bacterium [16, 43, 50, 100, 113, 175]. Fortunately, expression of Alg7 is not required in *E. coli* given a functional homolog, WecA, is already present in this host, having the inherent advantage of being a native enzyme as well as the ability to recognize the bacterial lipid carrier (undecaprenol or bactoprenol) as its natural substrate.

1.3.2 WecA

Analogous to Alg7, WecA also play a role in the initiation of the synthesis of certain bacterial glycans. WecA, also a GlcNAc-1-phosphate transferase, attaches

the first phospho-GlcNAc to bactoprenol, serving mainly as starting point for the synthesis of *O*-antigens and the enterobacterial common antigen [4]. As expected, WecA is also a membrane protein, having 11 transmembrane domains experimentally confirmed.

Given this features, it becomes obvious that WecA is the best candidate for initiating the synthesis of the trimannose core in *E. coli*. Also, given the prevalence of *O*-antigens and the enterobacterial common antigen on cells surface, we hypothesized that natural expression levels for WecA should also be enough for the production of adequate levels of the trimannose core.

1.3.3 Alg13 and Alg14

Addition of a second GlcNAc to the nascent glycan is catalyzed by a complex conformed by Alg13 and Alg14, where the active site is located on the Alg13 subunit and Alg14 serves as an anchor to the ER membrane [59]. Evidence also suggests that interaction between these two subunits is necessary for catalytic activity, given an Alg14 knock out compromises viability in *S. cerevisiae* indicating that Alg13 by itself can not catalyze the addition of GlcNAc in a proper way [57].

Alg13 is a soluble protein localized in the cytoplasm, which can form oligomers depending on its concentration. Its 3D structure has been solved and its sequence shows homology to the catalytic domain of bacterial MurG [177]. It has been also concluded that it is recruited by Alg14 to the cytoplasmic face of the ER membrane by specific hydrophobic interactions between its *N*-terminus and Alg14 *C*-terminus [57, 59].

On the other hand, Alg14 is a membrane protein that can be found an-

chored to the cytoplasmic face of the ER membrane. Its membrane topology has been established, concluding that it has 1 transmembrane domain with the *N*-terminus facing the ER lumen, and 2 membrane associated domains [8]. A complex conformed by Alg7, Alg13 and Alg14 has been reported [121], where Alg14 seems to be the central structure recruiting Alg13 from the cytoplasm and interacting with Alg7 thanks to one of its membrane associated domains and its *N*-terminal region [103]. The 3D structure of Alg14 has not being determined experimentally. However, thanks to its homology to the membrane domain of bacterial MurG and with help of the previously established structure of Alg13, a good approximation has been established based on computational structure analyses [103].

1.3.4 Alg1

Transfer of the first mannose residue to the early trimannose core is catalyzed by Alg1, a β1,4-mannosyltransferase. Its catalytic activity was identified 30 years ago [73]. However, our knowledge about this particular enzyme is still very limited. Alg1 is a membrane protein which proposed topology (not confirmed experimentally) consists of 4 transmembrane domains located towards the *N*-terminus and a soluble *C*-terminus facing the cytoplasm [58]. Notably, the *N*-terminus of Alg1 is not required for catalytic activity, and indeed, a soluble version of Alg1 lacking its transmembrane domain have been shown to be active *in vitro* [137].

Interestingly, in a simlar way to Alg14, Alg1 plays the role of central enzyme in a multiple unit complex comprised of Alg1, Alg2 and Alg11. Since it has been suggested that Alg2 and Alg11 do not interact with each other, Alg1 is most likely to be responsible for the complex assembly. Furthermore, Alg1 happens

to form dimers that later interact with Alg2 and Alg11 [58].

For the effects of this study, the most important feature about Alg1 is that it can be expressed recombinantly in *E. coli* [35]. Moreover, after lysing cells, activity of recombinant Alg1 was confirmed *in vitro* [137]. Although, *In vivo* activity in *E. coli* has not been shown before our study.

1.3.5 Alg2

The trimannose core structure is completed by addition of 2 mannose residues by Alg2, a dual function α1,3 and α1,6-mannosyltransferase [122]. As is the case for Alg1, the knowledge about Alg2 is still relatively limited, perhaps because it is an essential protein for many eukaryotes, constraining the possibilities for its analysis *in vivo*. Alg2 is also a membrane protein, which membrane topology has been experimentally confirmed. Studies suggest that it has 2 transmembrane domains at the *N*-terminus, a long soluble domain facing the cytoplasm, and 2 additional membrane-associated domains at the *C*-terminus [81]. It is also interesting that this glycosyltransferase does not require any of the two *N*-terminal transmembrane domains for its catalytic activity, while the same is true for the *C*-terminal membrane associated domains. However, deletion of all membrane domains seems to render the protein unstable [81].

As mentioned before, Alg2 is part of a complex conformed by Alg1, Alg2 and Alg11 [58] and its catalytic domain is most likely localized in the cytoplasmic soluble domain [81].

Finally, and of particular interest in this project, Alg2 has been expressed recombinantly in *E. coli* as a fusion to the *C*-terminus of Thioredoxin. Subsequent experiments demonstrated that membranes recovered from this *E. coli* strain

showed catalytic activity consistent with that expected for Alg2 [122]. Once again, *In vivo* activity in *E. coli* has not been shown before our study.

1.3.6 Sugar nucleotides: substrates for glycosyltransferases

Glycan synthesis requires more than just expression of glycosyltransferases. Carbohydrate subunits are provided by highly energetic compounds denominated nucleotide sugars. In the particular case of the trimannose core synthesis, only two sugar nucleotides are required: Uridine Diphosphate *N*-acetylglucosamine (UDP-GlcNAc) and Guanine Diphosphate Mannose (GDP-Man).

UDP-GlcNAc

UDP-GlcNAc is a substrate for both WecA and the Alg13-Alg14 complex. In *E. coli*, it is one of the precursors for peptidoglycan synthesis, so the endogenous pool of this sugar donor is considered to be enough for the engineering of glycan synthesis [141]. Also, the native synthesis pathway for regeneration of UDP-GlcNAc stocks has been shown to be efficient enough, so it should not be necessary to provide this precursor alongside with the growth medium [34].

GDP-Man

In *E. coli*, GDP-Man is synthesized *de novo* by two enzymes, ManB and ManC, using mannose-6-phosphate as substrate [160]. GDP-Man is then mainly used for synthesis of GDP-Fuc (Fuc: fucose) by two additional enzymes, GMD and WcaG, as part of the colanic acid synthesis pathway [44] hence GDP-Man does not accumulate in *E. coli* given its conversion into GDP-Fuc [160]. However,

there are several possible strategies for increasing the GDP-Man pool, like overexpression of RscA (a positive regulator for the colanic acid pathway), overexpression of ManB and ManC, and deletion of the *gmd* gene [141].

1.3.7 Synthesis of the trimannose core in *E. coli*

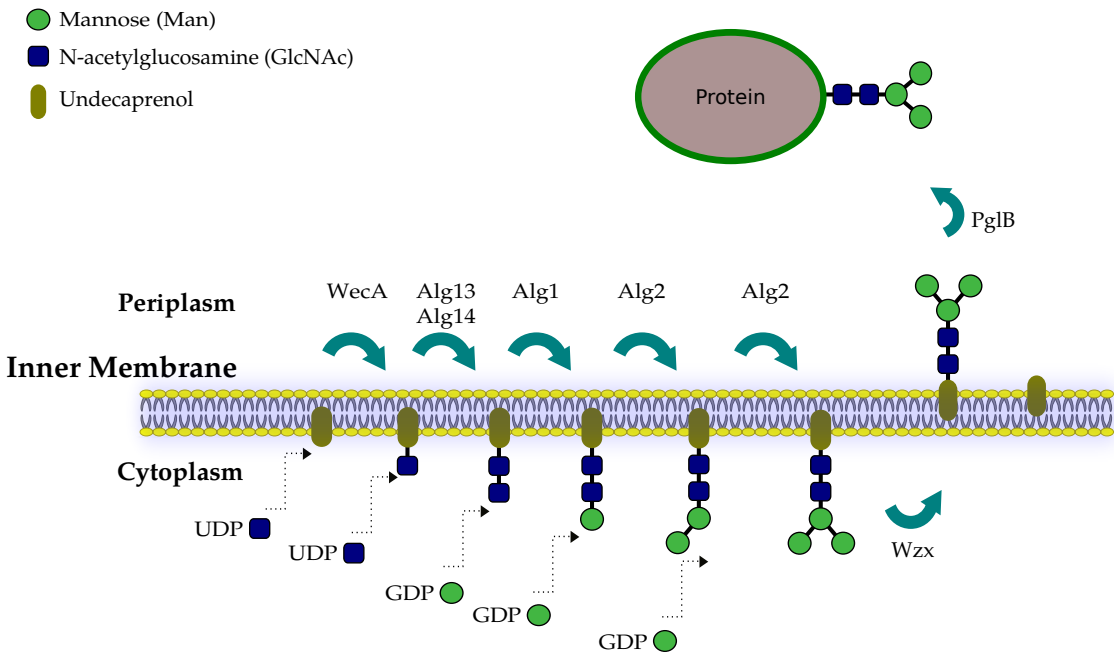

Figure 1.5: *Synthesis of the trimannose core in E. coli.* The highlighted residues constitute the "common core" or "trimannose core" (adapted from [115]). UDP: uridyl diphosphate, GDP: guanidyl diphosphate

The main hypothesis in this project is that the trimannose core can be synthesized in *E. coli* given the necessary glycosyltransferases are recombinantly expressed and the cytoplasmic concentration of required sugar nucleotides can be maintained at appropriate levels. After that, *C. jejuni* PglB would transfer the trimannose core to a target protein in the periplasm. Figure 1.5 shows a diagram summarizing this hypothesis.

The following chapters show how the synthesis of the trimannose core was

achieved, how it can be attached *in vivo* to model proteins, and some of the many directions this process may take respect to synthesis of complex glycans, optimization of the current process and alternative applications.

CHAPTER 2

AN ENGINEERED EUKARYOTIC PROTEIN GLYCOSYLATION PATHWAY IN *ESCHERICHIA COLI*[1]

2.1 Introduction

N-linked protein glycosylation is an essential and conserved process that occurs in the endoplasmic reticulum of eukaryotes [25]. It is the most common post-translational modification of eukaryotic proteins and can affect many important protein properties [67]. *N*-linked glycosylation is not limited to eukaryotes, however, as *bona fide N*-linked glycosylation pathways have been discovered in proteobacteria [162] and transferred to *E. coli*. There are several notable differences between bacterial and eukaryotic *N*-glycosylation systems. First, bacteria assemble oligosaccharides on undecaprenyl pyrophosphate (Und-PP) in the cytoplasmic membrane whereas eukaryotes use dolichyl pyrophosphate (Dol-PP) in the ER membrane. Second, the N-X-S/T consensus sequence for *N*-glycosylation in eukaryotes can be extended to D/E-X_{-1}-N-X_{+1}-S/T (X_{-1}, X_{+1} $\neq$ P) in bacteria [88]. Third, and most importantly, bacterial *N*-glycans [187] are completely distinct from any known eukaryotic glycan. As a result, glycoproteins derived from existing bacterial expression systems have been restricted to bioconjugate vaccines [51, 76] or glycoproteins that require extensive *in vitro* modification [147]. The construction of a eukaryotic glycosylation pathway in *E. coli* that generates human-like *N*-glycans has remained an elusive challenge despite years of speculation [30, 127, 178].

To address this challenge, we focused on engineering *E. coli* to produce mannose$_3$-*N*-acetylglucosamine$_2$ ($Man_3GlcNAc_2$) eukaryotic glycans. We chose

[1]Adapted with permission from Valderrama-Rincon, J. D. et al. An engineered eukaryotic protein glycosylation pathway in Escherichia coli. Nature Chemical Biology (2012).

$Man_3GlcNAc_2$ because it is: (i) the core structure common to all human *N*-glycans; (ii) the predominant *N*-glycan produced by baculovirus-insect cells [89], carrot root plant cells [9], and *Tetrahymena thermophila* [179], all of which yield glycans that are fit for pre-clinical and clinical products; and (iii) the minimal glycan required for a therapeutic glycoprotein currently on the market [170]. To generate $Man_3GlcNAc_2$ glycans, a synthetic pathway was designed that we predicted would assemble the glycan on Und-PP in the cytoplasmic membrane of *E. coli* (figure 2.1a). The first step in this pathway involved the endogenous integral membrane protein WecA, a glycosyltransferase (GTase) that catalyzes the transfer of GlcNAc-1-phosphate to undecaprenyl phosphate (Und-P). The native GlcNAc-PP-Und intermediate served as the foundation for the $Man_3GlcNAc_2$ structure.

2.2 Results and discussion

To extend the glycan, several heterologous GTases from *Saccharomyces cerevisiae* were selected because these have been solubly expressed in E. coli [35, 122, 177] and in some cases the expressed enzymes were active *in vitro* [35, 122]. Specifically, for addition of the second GlcNAc residue to GlcNAc-PP-Und, we chose the *S. cerevisiae* β1,4-GlcNAc transferase that is comprised of the subunits Alg13 and Alg14. In yeast, Alg14 is an integral membrane protein that functions as a membrane anchor to recruit soluble Alg13 to the cytosolic face of the ER membrane, where synthesis of GlcNAc2-PP-Dol occurs [18]. For the subsequent steps, we employed *S. cerevisiae* β1,4-mannosyltransferase Alg1, which specifies the addition of the first mannose to the glycan [35], and the bifunctional mannosyltransferase Alg2, which carries out the subsequent addition of both an α1,3-

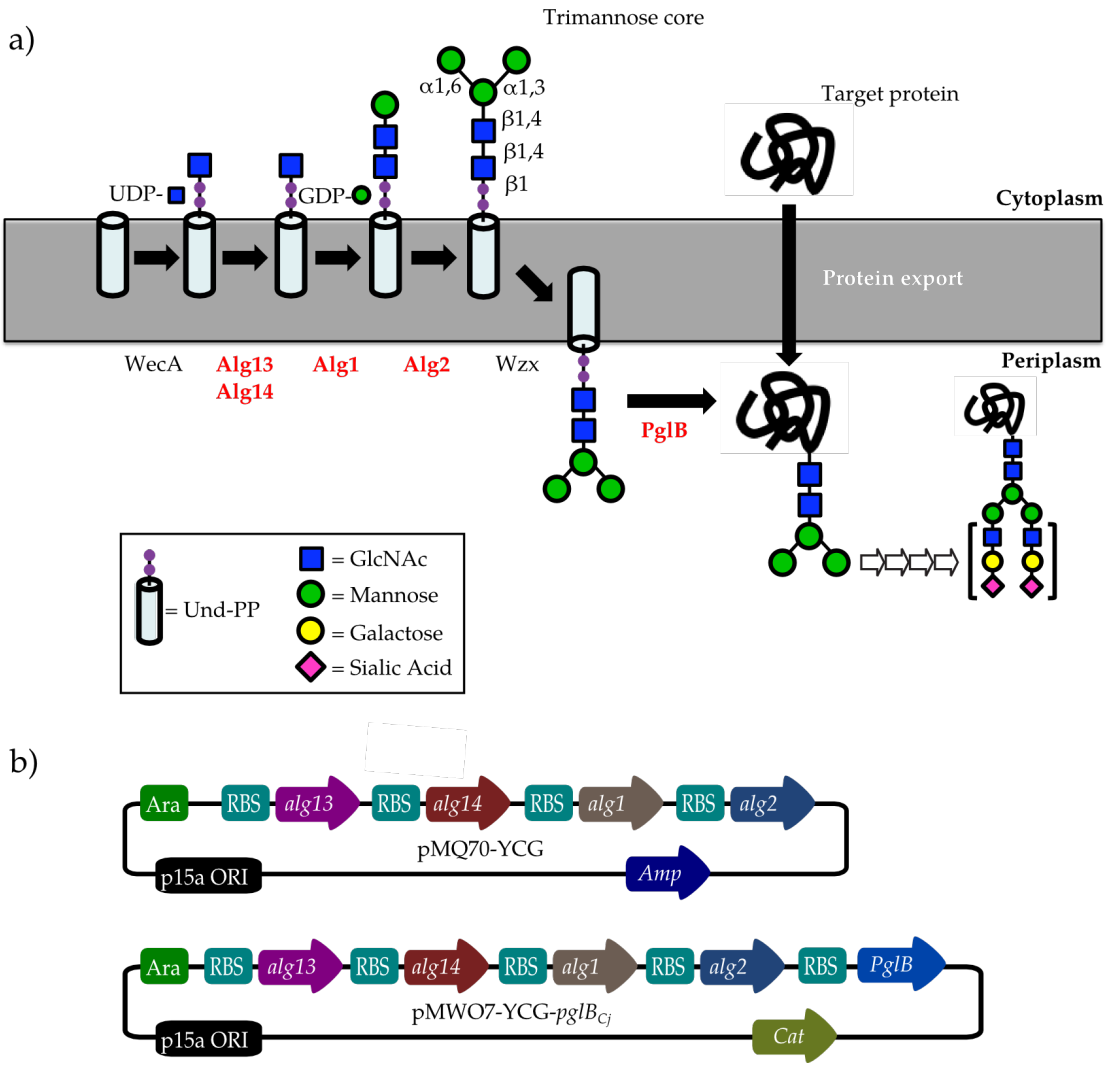

Figure 2.1: *Engineering eukaryotic glycan biosynthesis in E. coli.* (a) Schematic of the synthetic pathway for synthesis of a trimannosyl core glycan and transfer to acceptor sites in target proteins. Enzyme names in black are native to *E. coli*; enzyme names in red are heterologous. See text for details. Glycan in brackets to the right of open arrows depicts terminally sialylated structure common in human glycoproteins, but outside the scope of this study. (b) Plasmid pMQ70-YCG for the biosynthesis of a trimannosyl core glycan and pMWO7-YCG-*pglB*$_{Cj}$ for producing a trimannosyl core glycan with the *C. jejuni* OTase PglB, were constructed in plasmids pMQ70 and pMW07, respectively.

and α1,6-mannose in a branched configuration [122]. Assembly of a functional $Man_3GlcNAc_2$ synthesis pathway required expression and localization of these enzymes in the inner membrane of *E. coli*. This represented a non-trivial task

given that the functional expression of even one non-enzyme eukaryotic membrane protein in *E. coli* can pose a significant challenge [85]. When co-expressed in *E. coli*, both Alg 13 and Alg14 were detected in the membrane fraction while Alg13 was also evident in the soluble fraction (figure 2.2), consistent with their localization in yeast. Likewise, both Alg1 and Alg2 were detected in the membrane fraction (figure 2.2).

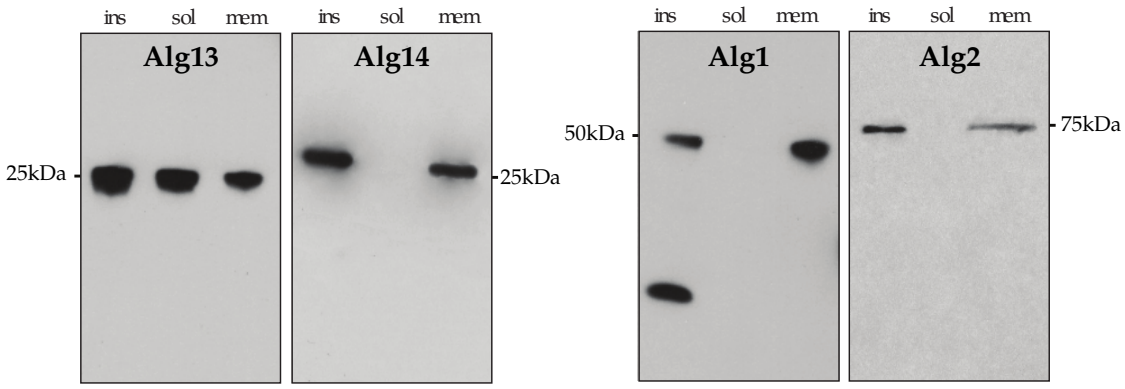

Figure 2.2: *Expresion of S. cerevisiae Alg13, Alg14, Alg1 and Alg2 in E. coli*. Western blot analysis of insoluble (ins), soluble (sol) and membrane (mem) fractions isolated from MC4100 *gmd::kan* cells carrying pMQ70-YCG. All proteins migrated according to their predicted molecular weights. Alg14 was detected using anti-FLAG antibodies; the rest were detected using anti-His antibodies as described in the Supplemental Methods.

To determine if the membrane-localized Alg enzymes were oriented correctly with their active sites on the cytoplasmic face of the inner membrane and expressed as active enzymes capable of producing $Man_3GlcNAc_2$ on Und-PP, we constructed plasmid pMQ70-YCG that encoded a synthetic gene cluster comprised of *alg13*, *alg14*, *alg1* and *alg2* (figure 2.1 1b). To increase the availability of the GDP-mannose substrate for Alg1 and Alg2, the gene encoding GDP-mannose dehydratase was deleted from *E. coli* strain MC4100. This deletion prevents the conversion of GDP-mannose to GDP-4-keto-6-deoxymannose in the first step of GDP-L-fucose synthesis [141]. To assay glycan synthesis, we exploited the fact that bacterial cell surfaces can display engineered oligosac-

charide structures in their lipopolysaccharide layer [52, 77, 186]. This approach is dependent upon the *O*-antigen ligase WaaL, which catalyzes the transfer of Und-PP-linked oligosaccharides to the *E. coli* lipid A core. These oligosaccharides are then shuttled to the cell surface where they can be conveniently labeled with fluorescent probes [52, 77].

Upon labeling with fluorescent *Canavalia ensiformis* concanavalin A (ConA), a lectin that binds terminal α-mannose, MC4100 *gmd::kan* cells expressing the synthetic pathway but not empty vector control cells became highly fluorescent (figure 2.3a). The fluorescence was clearly localized on the cell surface (figure 2.4). Importantly, cell fluorescence was significantly diminished in the absence of *alg1* or *alg2* (figure 2.3a), confirming that terminal α-mannose residues on the cell surface were dependent on these enzymes. Likewise, when the synthetic pathway was expressed in MC4100 *gmd::kan* that also lacked *waaL*, cells were minimally fluorescent (figure 2.3a). Taken together this analysis suggests that oligosaccharides with terminal α-mannose residues were synthesized and shuttled to the cell surface via the WaaL enzyme. As an important corollary, a native *E. coli* flippase (e.g., Wzx) must be involved since WaaL-dependent transfer to lipid A requires that Und-PP-linked oligosaccharides are available on the periplasmic face of the cytoplasmic membrane [5].

To verify the structure of the glycans, lipid-linked oligosaccharides (LLOs) were extracted and subjected to acid hydrolysis to remove glycans. The released glycans were permethylated and characterized by matrix-assisted laser desorption/ionization tandem time-of-flight (MALDI-TOF/TOF) analysis. The MALDI-MS spectrum revealed $Hex_3HexNAc_2$ as the primary oligosaccharide, which is consistent with the expected $Man_3GlcNAc_2$ glycan. In addition, $Hex_2HexNAc_2$ and $Hex_4HexNAc_2$ oligosaccharides were detected (figure 2.3b,

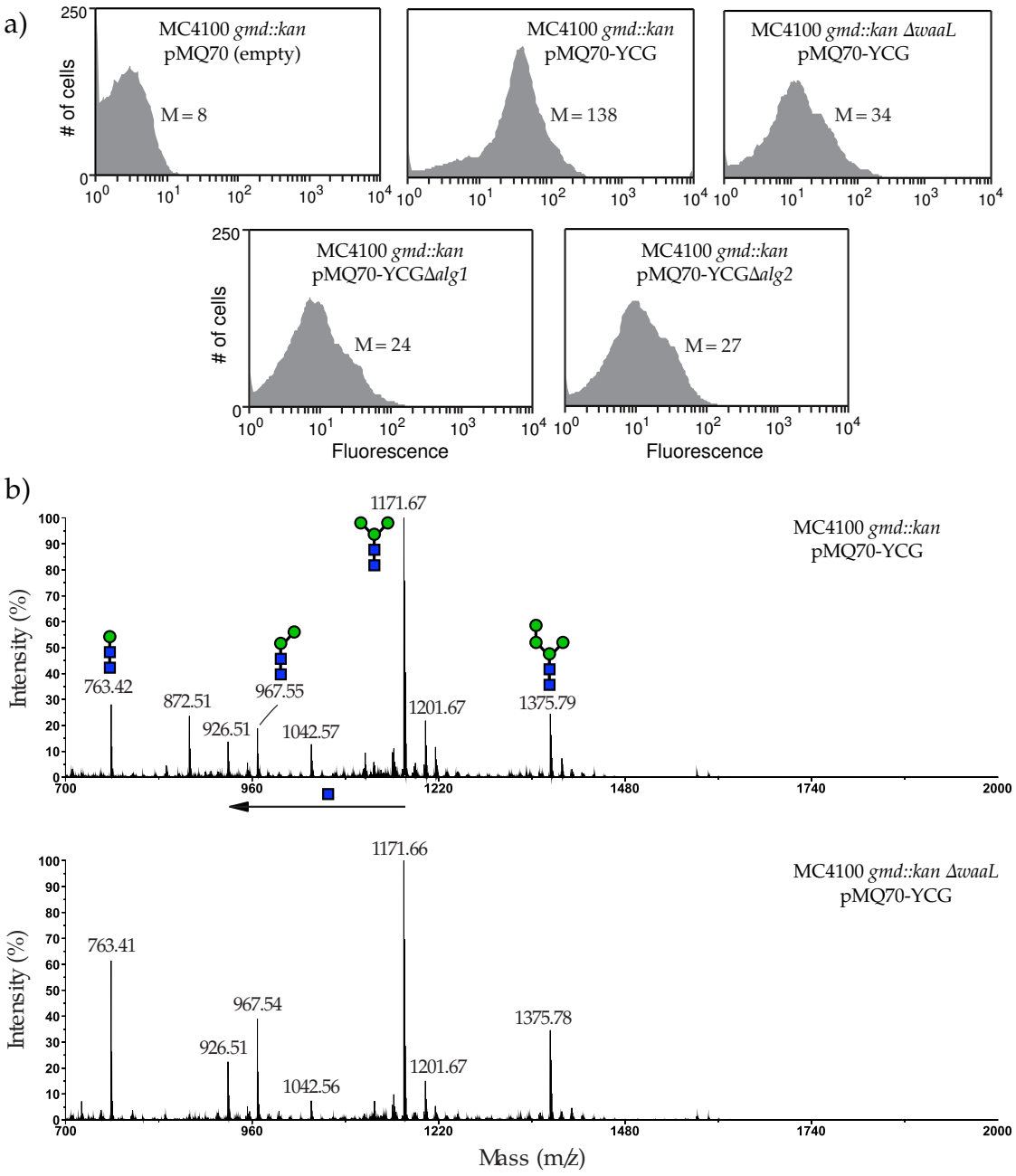

Figure 2.3: *Characterization of LLOs produced by glycoengineered E. coli.* (a) Flow cytometric analysis of E. coli MC4100 *gmd::kan* or MC4100 *gmd::kan ΔwaaL* cells carrying plasmids as indicated. Cells were labeled with ConA-AlexaFluor prior to flow cytometry. Median cell fluorescence (M) values are given for each histogram. (b) MALDI-MS profile of permethylated glycans released from LLOs by acid hydrolysis. LLOs were extracted from E. coli MC4100 *gmd::kan* (top panel) or MC4100 *gmd::kan ΔwaaL* cells (bottom panel) carrying plasmid pMQ70-YCG. The major signal at *m/z* 1171 corresponds to $[M+Na]^+$ of $Hex_3HexNAc_2$.

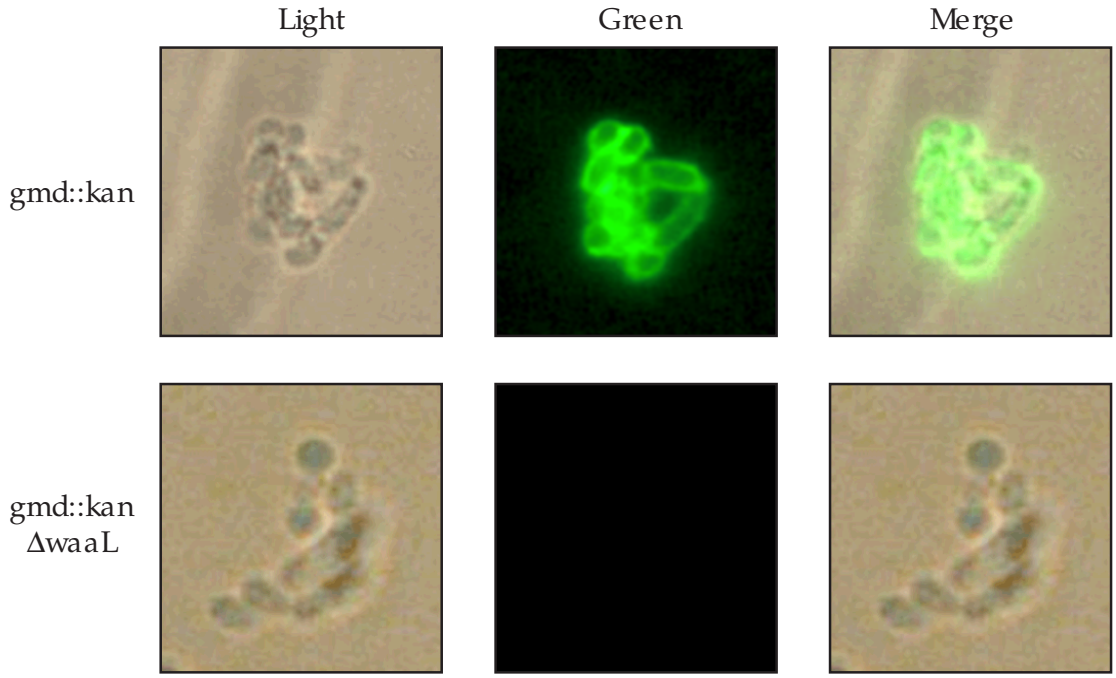

Figure 2.4: *ConA-AlexaFluor labeling*. ConA-AlexaFluor labeling of MC4100 *gmd::kan* (top panels) or MC4100 *gmd::kan ΔwaaL* (bottom panels) cells each carrying plasmid pMQ70-YCG. Cells were visualized by light and fluorescence microscopy as described in the Materials and Methods section. Panels show phase contrast microscopy (left), fluorescence microscopy using green emission filter (center) and merge (right).

top panel). The MALDI-MS spectrum of LLOs isolated from the MC4100 *gmd::kan ΔwaaL* strain was also consistent with the presence of $Man_3GlcNAc_2$ (figure 2.3b, bottom panel). This confirmed that the lack of cell surface labeling observed for these cells was a result of the *waaL* deletion and not the inability to synthesize oligosaccharides. Finally, released glycans analyzed by ^1H NMR spectroscopy were consistent with the eukaryotic core glycan Manα1-3(Manα1-6)-Manβ1-4-GlcNAcβ1-4-GlcNAc (figure 2.5 and 2.6).

We next investigated whether $Man_3GlcNAc_2$ glycans could be transferred to secretory glycoproteins by a heterologous oligosaccharyltransferase (OTase) *in vivo*. To our knowledge, no eukaryotic OTases have been functionally expressed in *E. coli*. Therefore, we focused our attention on PglB from *C. jejuni*

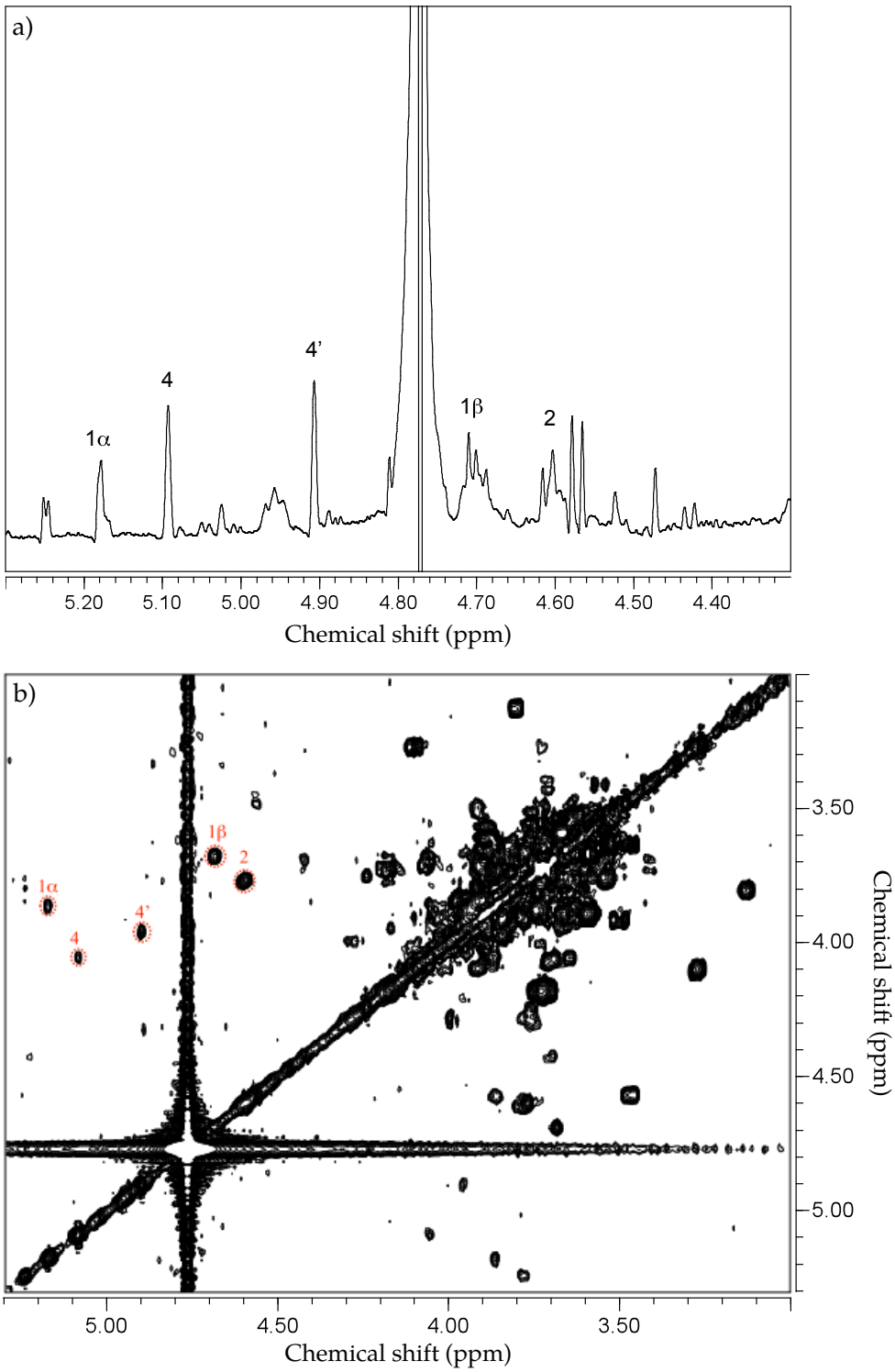

Figure 2.5: *^{1}H NMR analysis of glycans released from extracted LLOs.*. (a) Anomeric region of the resolution-enhanced 1-D proton spectrum acquired at 25°C. (b) Partial 2-D gCOSY spectrum acquired at 25°C. This analysis confirms the identity of every single carbohydrate unit in the glycan as well as the nature of the linkages between them.

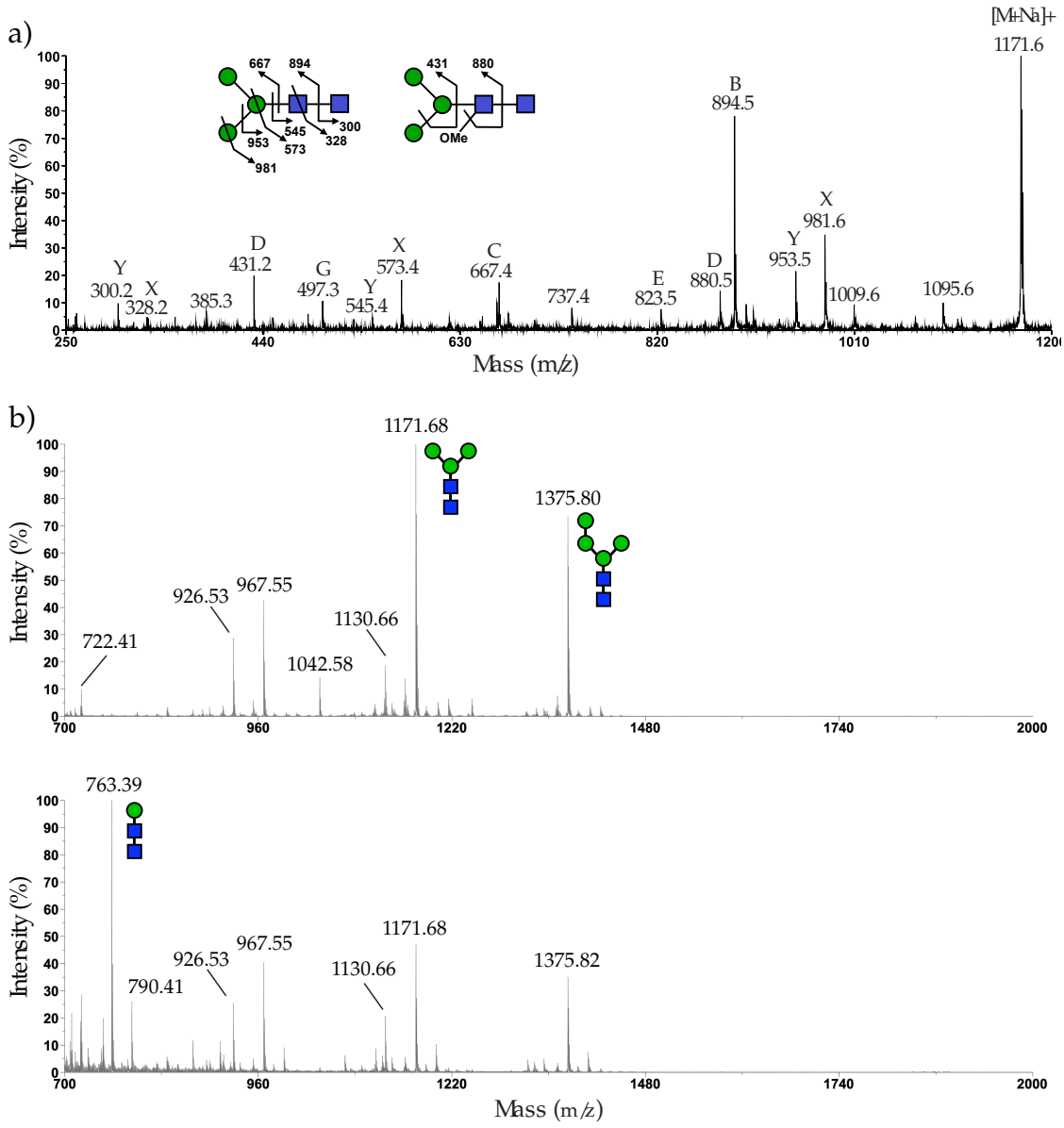

Figure 2.6: *Mass spectrometry analysis of glycans released from LLOs and glycoprotein.* (a) MALDI-TOF/TOF MS^2 sequencing of *N*-glycans at *m/z* 1171. X/Y ion series at *m/z* 300/328, 545/573, 953/981, coupled with B/Y ion pair at *m/z* 894/300, showed the presence of biantennary trimannosyl-core *N*-glycan structure. D ions at *m/z* 431 and 880 further confirmed the structure. (b) MALDI-MS profile of glycans released from scFv13-$R4^{1x-DQNAT}$ by PNGaseF (top panel) and treated further with *C. ensiformis* α-exomannosidase (bottom panel) to determine the identity of terminal hexose units. This confirms that the two terminal mannose residues were α-linked to the first mannose residue.

(PglB$_{Cj}$) because it is the best characterized bacterial OTase [102] and is capable of handling diverse Und-PP-linked oligosaccharides as substrates in *E. coli* [51, 76, 172]. For secretory glycoprotein targets, we initially examined (i) *E. coli* maltose binding protein (MBP) which is a native periplasmic protein and (ii) anti-β-galactosidase single-chain antibody fragment called scFv13-R4 that was modified with an *N*-terminal co-translational export signal from *E. coli* DsbA [144]. Each of these target proteins was modified at the *C*-terminus with four tandem repeats of the bacterial glycan acceptor motif DQNAT [52]. MC4100 *gmd::kan ΔwaaL* cells were transformed with plasmids encoding one of these target proteins, the Man$_3$GlcNAc$_2$ synthesis pathway, and PglB$_{Cj}$. Following co-expression, a low but consistent amount of MBP$^{4x-\mathrm{DQNAT}}$ and scFv13-R4$^{4x-\mathrm{DQNAT}}$ was bound by ConA, in contrast to the same target proteins produced in cells carrying an inactive PglB$_{Cj}$ mutant [173] (figure 2.7a). Further, when target proteins from cells with wildtype PglB$_{Cj}$ were treated with peptide:*N*-glycosidase F (PNGase F), an amidase that specifically cleaves between a reducing-end GlcNAc and asparagine, binding by ConA was eliminated (figure 2.7a). Thus, α-mannose-containing glycans were linked specifically to asparagines in the target proteins by PglB$_{Cj}$. PNGase F-released glycans from glycosylated scFv13-R4$^{4x-\mathrm{DQNAT}}$ were characterized by MALDI-TOF/TOF analysis and the predominant *N*-linked glycan was Hex$_3$HexNAc$_2$ along with a lesser amount of Hex$_4$HexNAc$_2$ (figure 2.7b). In addition, a version of scFv13-R4 that carried a single *C*-terminal DQNAT sequon was subjected to nonspecific proteolytic digestion with Pronase E and permethylation prior to MS analysis [101]. The major ion seen at *m/z* 1282 was consistent with Man$_3$GlcNAc$_2$-Asn, wherein the asparagine residue underwent β-elimination during the permethylation procedure (figure 2.7c) [101]. MS^2 sequencing of the glycan at *m/z*

1171 confirmed the biantennary trihexosyl structure (figure 2.8a) and ^{1}H NMR analysis on PNGase F-released glycans was consistent with Manα1-3(Manα1-6)-Manβ1-4-GlcNAcβ1-4-GlcNAc (figures 2.5 and 2.9).

The increased abundance of the larger glycoform in the protein samples relative to the LLO samples may have been due to differences in the culture conditions and/or enrichment of higher mannose glycans during ConA affinity chromatography. When PNGase F-released glycans were treated with *Canavalia ensiformis* α-exomannosidase to specifically hydrolyze terminal α-mannose residues, the MALDI-MS spectrum showed the emergence of $HexHexNAc_2$ as the predominant glycoform at the expense of both $Hex_3HexNAc_2$ and $Hex_4HexNAc_2$ (figure 2.8b). Consistent with this MS spectrum, NMR analysis revealed a residue with H-1 (5.080 ppm) and H-2 (4.065 ppm) chemical shifts consistent with an additional Man linked to one of the branching Man residues (Fig. S3). The presence of a putative $Man_4GlcNAc_2$ is somewhat surprising because Alg2 is not known to specify the addition of mannose beyond $Man_3GlcNAc_2$-PP-Dol *in vitro* [122]. Elongation of $Man_3GlcNAc_2$ to $Man_4GlcNAc_2$ and $Man_5GlcNAc_2$ is attributed to the bifunctional Alg11 enzyme. It should be noted, however, that both $Man_3GlcNAc_2$-PP-Dol and $Man_4GlcNAc_2$-PP-Dol accumulated in a *S. cerevisiae alg11* mutant [33], suggesting that Alg1 or Alg2 may catalyze $Man_4GlcNAc_2$-PP-Dol production.

To determine whether $Man_3GlcNAc_2$ glycans could be transferred to eukaryotic glycoproteins, we next attempted to glycosylate: (i) the Fc domain of human IgG1 at its conserved N297 glycosylation site, (ii) bovine ribonuclease A (RNaseA) at its N34 acceptor site, and (iii) the placental variant of human growth hormone (hGHv) at its N140 glycosylation site. The open reading frames encoding these proteins were cloned downstream of an *N*-terminal

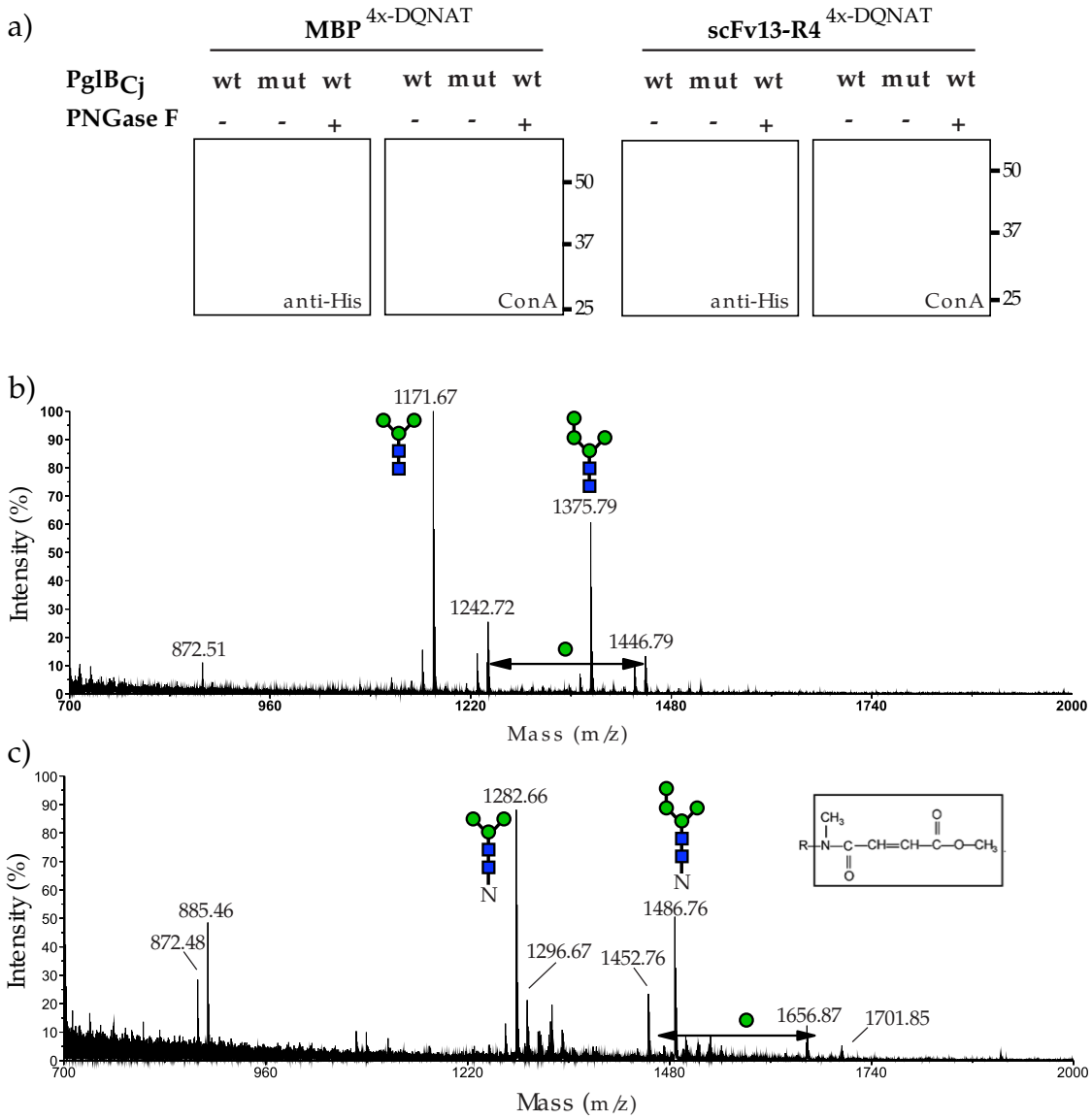

Figure 2.7: *Transfer of eukaryotic glycans to target proteins in E. coli.* (a) Western blot analysis of MBP$^{4x-DQNAT}$ and scFv13-R4$^{4x-DQNAT}$ affinity purified from E. coli MC4100 *gmd::kan ΔwaaL* cells carrying pMWO7-YCG-PglB$_{Cj}$ or pMWO7-YCG-PglB$_{Cjmut}$ as indicated. Proteins isolated from cells expressing wild-type PglB$_{Cj}$ were further treated with PNGase F for removal of *N*-linked glycans. Polyhistidine tags on the proteins were detected using anti-His antibodies while mannose glycans on the proteins were detected using ConA. (b) MALDI-MS profile of permethylated glycans released from scFv13-R4$^{4x-DQNAT}$ by PNGase F treatment. The major signal at *m/z* 1171 corresponds to [M+Na]$^+$ of $Hex_3HexNAc_2$. (c) MALDI-MS profile of permethylated glycopeptides generated by digestion of scFv13-R4$^{1x-DQNAT}$ with Pronase E. The major signal at *m/z* 1282 corresponds to the permethylation product of $Hex_3HexNAc_2$-N, where the asparagine residue underwent β-elimination during the permethylation procedure (see inset).

Partial assignmet of the NMR signals

No.	Residue	Position	Chemical Shift (ppm)[a]						NOE
			1	2	3	4	5	6	
I	4-α-GlcNAc	1	5.183	3.88	3.65	n.d.	n.d.	n.d.	
II	α-Man	4	5.094	4.059	3.89	n.d.	n.d.	n.d.	VI-3
III	?	?	5.080	4.065	n.d.	n.d.	n.d.	n.d.	
IV	?	?	5.042	4.065	n.d.	n.d.	n.d.	n.d.	
V	α-Man	4′	4.910	3.97	3.89	n.d.	n.d.	n.d.	
VI	3,6-β-Man	3	4.777	4.25	3.75	n.d.	n.d.	n.d.	
VII	4-β-GlcNAc	1	4.692	3.69	n.d.	n.d.	n.d.	n.d.	
VIII	4-β-GlcNAc	2	4.602	3.79	3.62	n.d.	n.d.	n.d.	
IX	Xyl	*	4.473	3.29	3.55	3.78	4.10/3.37		

Figure 2.8: *Partial assignment of the NMR signals belonging to $Man_3GlcNAc_2$ from protein-released glycans*. The chemical shifts obtained from the 2-D spectra were consistent with the main component of the sample being $Man_3GlcNAc_2$. This sample was available in sufficient quantity to allow acquisition of a 2-D HSQC spectrum, which confirmed the proton peak assignments. Xylose was determined to be a contaminant as it was not detected by MS analysis or by NMR of glycans released from lipids. Xylose is a common contaminant of bacterial samples.

DsbA export signal or full-lenth MBP in the case of hGHv. Since the N-X-S/T consensus motif in eukaryotes is extended to D/E-X_{-1}-N-X_{+1}-S/T in bacteria [88], we used Fc and hGHv mutants whose native glycosylation motifs were mutated from QYNST (residues 295-299) and IFNQS (residues 138-142), respectively, to DQNAT. Likewise, we used an RNaseA variant with an S32D substitution [87]. Expression of Fc^{DQNAT}, $hGHv^{DQNAT}$, or RNaseA S32D in the periplasm of *E. coli* cells carrying the pMWO7-YCG-$PglB_{Cj}$ plasmid yielded clearly glycosylated proteins (figure 2.10a and b). Whereas glycosylation of the structurally flexible acceptor sites in the Fc domain and hGHv was not surprising in light of previous studies [52, 147], RNaseA glycosylation was unexpected given that the acceptor site is located in a structured domain that is not glycosylated by $PglB_{Cj}$ *in vitro* [87]. These data imply that $PglB_{Cj}$ may be capable of glycosylating residues in both unstructured and structured regions of eukaryotic acceptor

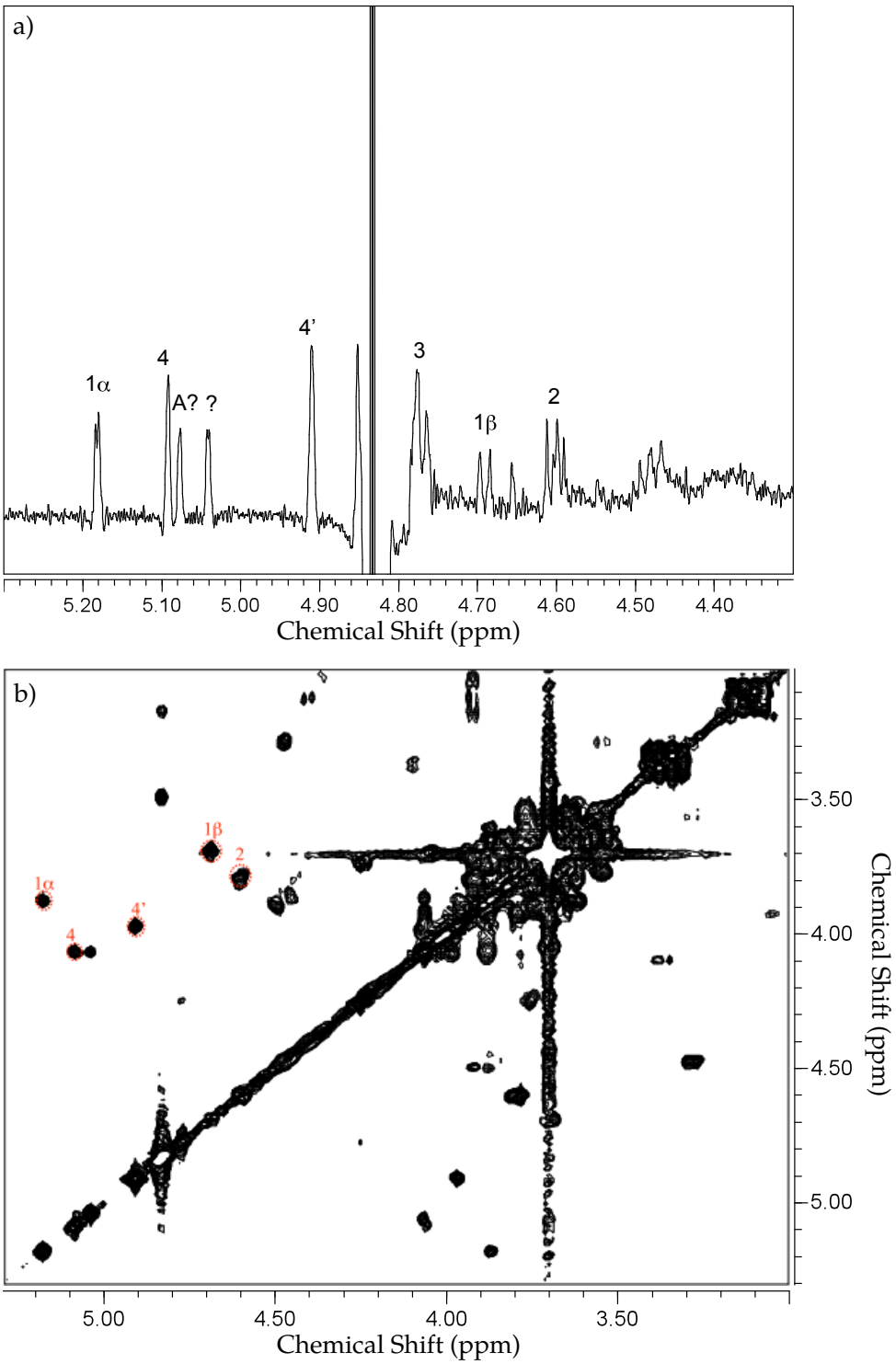

Figure 2.9: *^{1}H NMR analysis of N-linked glycans released from purified scFv13-R4$^{4x-DQNAT}$.* (a) Anomeric region of the resolution-enhanced 1-D proton spectrum acquired at 21°C. (b) Partial 2-D gCOSY spectrum acquired at 21°C. This analysis confirms the identity of every single carbohydrate unit in the glycan as well as the nature of the linkages between them.

proteins *in vivo*.

In this work, we have established glycosylation of target proteins with $Man_3GlcNAc_2$ glycans in recombinant *E. coli*. This was made possible by the functional co-expression of four membrane-bound eukaryotic enzymes and one bacterial integral membrane protein. The significance of this result is perhaps best reflected by the well-documented difficulties in expressing just one eukaryotic membrane protein in *E. coli* [85]. The further requirement that the expressed membrane proteins must be catalytically active would not have been predicted from *in vitro* studies alone. For example, earlier *in vitro* studies reported that $PglB_{Cj}$ was not capable of utilizing chitobiose donors such as dolichyl-pyrophosphate-GlcNAc-GlcNAc disaccharide [29]. Also, $PglB_{Cj}$ was hardly able to glycosylate RNaseA *in vitro* [87]. However, under the *in vivo* conditions tested here, $PglB_{Cj}$ was capable of transferring lipid-linked pyrophosphate-GlcNAc-GlcNAc saccharide substrates to both unstructured and structured regions of diverse targets including RNaseA.

Despite our success in reconstituting a eukaryotic glycosylation pathway in *E. coli*, there remain some important challenges that will need to be overcome for the practical application of this technology. For example, only a small fraction ($<$1%) of each expressed protein was glycosylated under the conditions tested here. With that said, the yield of glycosylated proteins has reached up to 50 μg/L in our hands and might be further improved by increasing expression in the periplasm, relieving enzymatic and metabolic bottlenecks, and/or optimizing the glycosylation enzymes. Along these lines, simple optimization strategies have previously been used to generate nearly 25 mg/L of bacterial glycoproteins in E. coli [76]. We anticipate further improvements will be achieved by applying new glyco-display technologies in bacteria, including cell surface and

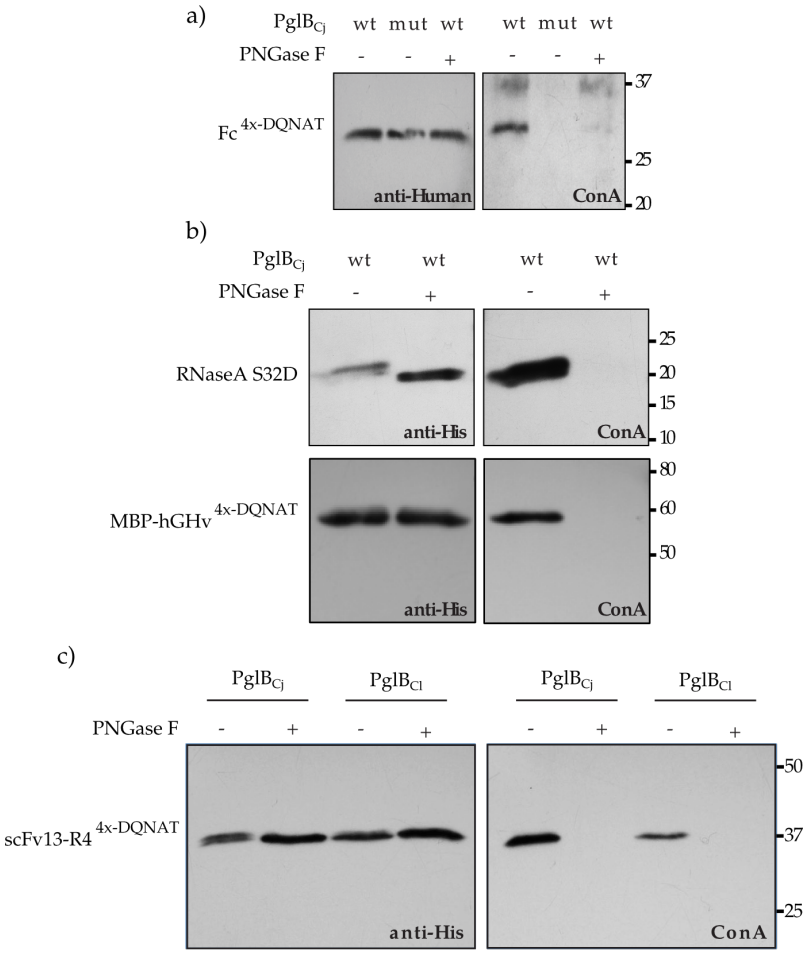

Figure 2.10: *Glycosylation of genuine human glycoproteins.* (a) Western blot analysis of Fc^{DQNAT} affinity purified from *E. coli* MC4100 *gmd::kan ΔwaaL* cells carrying pMWO7-YCG-PglB$_{Cj}$ or pMWO7-YCG-PglB$_{Cjmut}$ as indicated. Proteins isolated from cells expressing wild-type PglB were further treated with PNGase F for removal of *N*-linked glycans. Fc^{DQNAT} was detected using anti-Human antibodies while mannose glycans on the proteins were detected using ConA. (b) Western blot analysis of RNaseA S32D and MBP-hHGv^{DQNAT} affinity purified from *E. coli* MC4100 *gmd::kan ΔwaaL* cells carrying pMWO7-YCG-PglB$_{Cj}$. Proteins were loaded directly or further treated with PNGase F for removal of *N*-linked glycans. RNaseA S32D and MBP-hHGv^{DQNAT} were each detected using anti-His antibodies while terminal mannoses on protein glycans were detected using ConA. (c) Transfer of trimannosyl glycans to target proteins by *C. lari* PglB. Western blot analysis of scFv13-R4$^{4x-DQNAT}$ affinity purified from *E. coli* MC4100 *gmd::kan ΔwaaL* cells carrying pMWO7-YCG-PglB$_{Cj}$ for expressing *C. jejuni* PglB or pMWO7-YCG-PglB$_{Cl}$ for expressing *C. lari* PglB. Proteins were loaded directly or further treated with PNGase F for removal of *N*-linked glycans. scFv13-R4$^{4x-DQNAT}$ was detected using anti-His antibodies while terminal mannoses on protein glycans were detected using ConA. In general, these results show that glycosylation of genuine human glycoproteins is possible, being this system a potential platform for production of therapeutic glycoproteins.

phage display systems [27, 48, 52] for high-throughput screening of glycosylation phenotypes. We also believe that other OTases, such as *C. lari* PglB ($PglB_{Cl}$) which is capable of transferring $Man_3GlcNAc_2$ (figure 2.10c) and can transfer glycans to minimal N-X-S/T acceptor sites [148], can be optimized for improved function. Alternatively, single-subunit eukaryotic OTases (e.g., *Leishmania major* STT3 paralogues [128]) could be used that naturally use as substrates both (i) eukaryotic glycoproteins with minimal N-X-S/T acceptor sites and (ii) eukaryotic glycans comprised of the trimannosyl core structure.

Since it does not have native glycosylation pathways, *E. coli* is the only platform for glycoprotein expression that offers bottom-up synthesis of novel glycan structures by expression of diverse GTases and OTases. The engineering of defined glycosylation pathways in *E. coli* sets the stage for further engineering of this host for the production of vaccines and therapeutics with even more structurally complex human-like glycans. Moreover, we believe glycoengineered *E. coli* has the potential to serve as a model genetic system for deciphering the glycosylation code which governs the non-template driven synthesis of diverse glycans and their specific attachment to proteins.

2.3 Materials and methods

2.3.1 Bacterial strains and media

Antibiotic selection was maintained at: 100 μg/mL ampicillin (Amp), 25 μg/mL chloramphenicol (Cam) and 50 μg/mL kanamycin (Kan). Luria Bertani (LB) media was used for *E. coli*, supplemented with glucose at 0.2% as indicated. Protein expression was induced by adding L-arabinose and isopropyl

-d-thiogalactoside (IPTG) at 0.2% and 100 mM, respectively. Yeast FY834 was maintained on YPD media and synthetic-defined-Uracil media was used to select or maintain yeast plasmids. *E. coli* MC4100 (F- araD139 Δ(argF-lac)U169 flbB5301 deoC1 ptsF25 relA1 rbsR22 rpsL150 thiA) was used as the recipient strain for genetic manipulations. To delete the *gmd* gene in MC4100 cells, P1*vir* phage transduction was performed using standard methods and the Keio collection [10]. The Kan resistance cassette was removed from MC4100 *waaL::kan* cells using plasmid pCP20 [36] prior to P1*vir* phage transduction to obtain the strain MC4100 *gmd::kan ΔwaaL*.

2.3.2 Plasmid construction

Plasmid pTrc99A-*alg13::alg14* was used for co-expression of yeast Alg13 and Alg14. Plasmid pTrc99A-*alg1* and pBAD(*alg2*)-DEST493 were used for individual expression of Alg1 and Alg2, respectively. pBAD(*alg2*)-DEST49 was a generous gift from Dr. Barbara Imperiali, while pTrc99A-*alg13::alg14* and pTrc99A-*alg1* were constructed using standard techniques. *alg13*, *alg14* and *alg1* genes were amplified from *Saccharomyces cerevisiae* genomic DNA. A *C*-terminal 6x-His tag was added to Alg1 and Alg13 while an *N*-terminal FLAG epitope tag was added to Alg14. An additional ribosomal binding site was introduced in front of *alg14* to allow bicistronic expression. Plasmids pTrc99A-MBP$^{4x-DQNAT}$, pTrc99A-ssDsbA-scFv13-R4$^{4x-DQNAT}$ and pTrc-ssDsbA-Fc^{DQNAT} were used for expression of glycoproteins, as reported previously [52]. Plasmid pTrc99A-ssDsbA-scFv13-R4$^{4x-DQNAT}$ was constructed as described for pTrc99A-ssDsbA-scFv13-R4$^{4x-DQNAT}$. The remaining plasmids were constructed using standard homologous recombination in *S. cerevisiae* as previously described [149]. Briefly, pMQ70-YCG (see figure 2.1) was constructed by first PCR amplifying the genes

alg13, *alg14*, *alg1*, and *alg2* from *S. cerevisiae* genomic DNA with primers containing appropriate regions of overlap. PCR products and the linearized vector pMQ70 were used to transform yeast strain FY834. Constructs assembled in yeast were electroporated into *E. coli* for verification via PCR, restriction enzyme digestion and/or sequencing. For glycosylation studies, vector pMW07 was similarly generated from pMQ70 by replacing the origin (ori) and antibiotic resistance cassette with the p15a ori and the cat gene for resistance to Cam. Plasmid pMW07 was the host vector for pMWO7-YCG-PglB$_{Cj}$ encoding *C. jejuni* PglB amplified from (i) pACYC-pgl (see figure 2.1) or (ii) a synthesized, codon optimized version of *C. jejuni* PglB (Mr. Gene), (iii) pMWO7-YCG-PglB$_{Cjmut}$ amplified from pACYC-pgl pglBmut encoding an inactive PglB variant [173] and (iv) pMWO7-YCG-PglB$_{Cl}$ encoding a codon optimized *C. lari* PglB synthesized for this work (Mr. Gene). The codon optimized version of PglB$_{Cj}$ showed a modest improvement over the wild-type enzyme (~15% increase in yield, data not shown) and was used to generate glycosylated scFv13-R4$^{1x-\mathrm{DQNAT}}$ for Pronase E digestion, scFv13-R4$^{4x-\mathrm{DQNAT}}$ for NMR analysis, RNAseA S32D for Western blot, and for the comparison of PglB$_{Cj}$ and PglB$_{Cl}$ (figure 2.6c). Plasmid pTrc99Y was generated from pTrc99A by adding the yeast origin and URA selection cassette from pMQ80 [149]. The gene encoding the bovine RNAseA S32D mutant was amplified from pMIK817 for pTrc99Y-ssDsbA-RNAseA S32D. The gene encoding hGHv was synthesized (Mr. Gene) and amplified for pTrc99Y-MBP-hGHv$^{\mathrm{DQNAT}}$. pMQ70-YCG*Δalg1* and pMQ70-YCG*Δalg2* were derived from pMQ70-YCG using standard digestion and ligation.

2.3.3 Isolation of membranes and LLOs

Membranes were isolated similarly to the method of Marani et al [106]. For LLOs extraction, overnight cultures were diluted 1:100 into LB broth and incubated at 37°C with shaking (250 rpm) until reaching A600 ~0.6. Protein expression was then induced with 0.2% *L*-arabinose and cells were incubated at 30°C overnight in screw-capped flasks filled with media. Lipid-linked oligosaccharides (LLOs) were extracted and partially purified as described previously [56]. Briefly, the LLO extraction protocolwas based on a total lipid extraction protocol using chloroform:water (2:1) as described previously [53]. This solvent mixture extracts the most hydrophobic lipids, leaving LLOs in the precipitate. The precipitate was washed with water to remove hydrophilic molecules and then LLOs were extracted preferentially using chloroform:methanol:water (10:10:3). Contaminating oligosaccharides not attached to lipids were removed by adsorbing extracted LLOs on DEAE-cellulose. LLOs were then recovered from DEAE-cellulose using an ammonium acetate solution. After releasing glycans from LLOs by acid hydrolysis, a butanol/water mixture was used to extract glycans to the water phase while lipids remained in the butanol phase. Finally, two ionic exchange resins (AG50W-X8 (Sigma) and AG1-X8 (BioRad)) were used for removing the remaining salts.

2.3.4 Flow cytometry

E. coli cells were incubated at 30°C overnight in LB supplemented with 0.2% L-arabinose in filled, sealed culture tubes. Cells were pelleted, washed, resuspended in (PBS) and boiled for 10 min. 2.5 μg/mL *Canavalia ensiformis* Concanavalin A (ConA)-AlexaFluor was added to the samples before incubation in

the dark for 15 min at room temperature. 100 μL of each sample were analyzed using a FACSCalibur (Becton Dickinson). Median fluorescence was determined from 10,000 events. For microscopy, 5 μL of cells with ConA-AlexaFluor were imaged on a Zeiss Axioskop 40 and all images were captured under bright field illumination or UV illumination.

2.3.5 Glycoprotein expression and purification

MC4100 *gmd::kan ΔwaaL* cells were freshly transformed with pYCG-PglB$_{Cj}$, and a plasmid for glycoprotein expression. Overnight cultures were diluted 1:100 into LB broth with antibiotics and incubated at 30°C with shaking (250 rpm) until reaching A600 ~3. Cultures were then induced with IPTG and L-arabinose, and returned to 30°C for ~16 h. Cells were pelleted, frozen, resuspended in ConA column buffer (50 mM HEPES, 0.15 M NaCl, 0.1 mM $CaCl_2$, 0.01 mM $MnCl_2$, pH 7.0) with 1.0 mg/mL lysozyme and disrupted by either sonication (for Western blot analysis) or with a single pass through a microfluidizer (Microfluidics #110Y) at 20,000 psi (for MALDI-MS analysis). To isolate glycoproteins, clarified lysates (10 min at 18,000xg at 4°C) were subject to affinity chromatography with a HiTrap ConA 4B column (GE Healthcare). Eluates were concentrated with 10-kDa MWCO columns (Sartorius Vivaspin20), diluted 40x in Ni-NTA buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, pH 8.0), and again concentrated to 1 mL. His-tagged glycoproteins were then purified with a Qiagen Ni-NTA Kit per manufacturers instructions. Fcs were then purified with a Nab Protein A/G Spin Kit (Thermo Scientific) per manufacturers instructions. For MALDI-TOF MS, eluate was dialyzed in a Slyde-A-Lyzer 3500 MWCO cassette (Thermo Scientific) in 1.8 L deionized water (24 h, three volume replacements). Proteins from negative control cells expressing PglB$_{Cjmut}$ were

processed identically and no glycoprotein could be detected by ConA affinity, as expected (data not shown). Aglycosylated PglB$_{Cjmut}$ control samples used in Western blot analysis were purified directly with a Qiagen Ni-NTA Kit or Nab Protein A/G Spin Kit per manufacturers instructions.

2.3.6 Western blot analysis

Purified protein samples were separated using 12% SDS-PAGE gels and transferred to PVDF membranes. Proteins that harbored 6x-His affinity tags were detected with a monoclonal anti-polyhistidine-horse radish peroxidase (HRP) conjugate (Sigma), a monoclonal anti-His (*C*-term) antibody (Invitrogen), or a monoclonal anti-FLAG® M2 antibodies (Stratagene), per manufacturers instructions. Human Fcs were detected by anti-Human HRP conjugate antibodies (Promega). Protein-conjugated glycans were detected with 2.5 μg/mL ConA-HRP conjugate (Sigma). To remove *N*-linked glycans prior to Western blotting, purified glycoproteins were incubated with 5,000 units of PNGase F (New England BioLabs) at 37°C for 1 hr. Samples that were not treated with PNGase F were handled identically except PNGase F was omitted from the solution.

2.3.7 Glycosylation characterization by mass spectrometry and NMR

Glycans were released from *E. coli* LLOs as described previously [56]. *N*-glycans were released from glycoproteins by treating 100 μg of purified scFv13-R4$^{4x-DQNAT}$ or scFv13-R4$^{1x-DQNAT}$ with trypsin followed by PNGase F digestion in 50 mM ammonium bicarbonate, pH 8. Glycans were collected as flow-

through after SepPak C18 (Waters) separation. Some glycans were further digested with *C. ensiformis* α-exomannosidase in 50 mM ammonium acetate, pH 4.5, followed by further desalting steps using a carbon column (envi-carb SPE tube, Supelco). The mannosidase treatment was halted prior to completion to fully examine the glycan intermediates. Permethylation of glycans was performed as described [79]. Asparagine-linked glycans were analyzed as described elsewhere [101] with permethylation. MALDI-MS profiles were acquired using a 4800 MALDI-TOF/TOF (Applied Biosystems) in positive ion reflectron mode (500-4000 Da mass range). MALDI-MS/MS sequencing of permethylated glycans was performed in positive ion reflectron mode. Air was used as the collision gas. Structure assignment was according to the method of Yu et al [188]. All NMR analysis was performed at the Complex Carbohydrate Research Center (University of Georgia, Athens, GA). Briefly, 1-D Proton, TOCSY and NOESY NMR spectra, run with water presaturation, and gradient enhanced COSY and HSQC spectra were acquired on a Varian Inova-600 MHz spectrometer, equipped with a cryoprobe, at 21°C (for protein-released glycans) and at 25°C (for lipid-released glycans). Chemical shifts were measured relative to internal acetone (δ_H = 2.218 ppm, δ_C = 33.0 ppm).

2.4 Contributions to this chapter

Juan D. Valderrama-Rincon and Adam C. Fisher designed research, performed research, analyzed data and wrote the paper. Judith H. Merritt designed research and performed research. Yao-Yun Fan performed MS analysis and analyzed data. Craig A. Reading and Krishan Chhiba performed research. Christian Heiss and Parastoo Azadi performed NMR analysis and analyzed data.

Markus Aebi designed research and analyzed data. Matthew P. DeLisa designed research, analyzed data and wrote the paper.

CHAPTER 3

FUNCTIONAL EXPRESSION OF ALG11, A GLYCOSYLTRANSFERASE FROM *SACCHAROMYCES CEREVISIAE*, IN *ESCHERICHIA COLI*

3.1 Introduction

As glycobiology is becoming an increasingly important field, the poor availability of affordable glycans has been a persistent limitation for several studies. While chemoenzymatic synthesis is the prevalent source of glycans for the field [19], bacterial *in vivo* synthesis has the potential to become a relatively inexpensive and versatile source for a variety of glycans. According to this, several strategies have been developed for production of diverse carbohydrates and glycans in engineered microorganisms [141]. However, given most glycosyltransferases are membrane proteins, their recombinant expression is often difficult [16, 43, 50, 100, 113, 175] and most of the time they require engineering of the host as well as engineering of the protein itself, like fusion to folding and membrane integration partners [72, 113, 120, 139, 184], before proper expression is achieved.

In general, it has been hypothesized that an ideal fusion partner for recombinant production of membrane proteins should be able to help the protein to insert properly in the membrane while bypassing the translocation machinery, so toxicity derived from translocon overloading would not be an issue [139]. One interesting example of such a partner is Mistic, a 13kDa protein from *Bacillus subtilis* which has been used successfully for overproduction of recombinant membrane proteins in bacteria [47, 139], constituting a promising fusion partner for glycosyltransferases.

Despite the challenges associated to expression of this glycosyltransferases,

therapeutic importance of glycans have encouraged a number efforts toward the synthesis of human-like glycans in several kinds of non-human cells more suitable for high scale production like Chinese hamster ovary cells, plant cells, insect cells and yeast cells. Synthesis of human-like glycans in these cells has been achieved mainly by recombinant expression of human glycosyltransferases [26, 69, 78, 180]. However, synthesis of such glycans in bacterial cells has not been reported until recently, when eukaryotic-like glycans were produced using *E. coli* as host [167]. In this case, the trimannose core, a common structure found on the vast majority of eukaryotic *N*-linked glycans [115, 146] was synthesized in *E. coli* by recombinant expression of several *S. cerevisiae* glycosyltransferases alongside with deletion of the *gmd* gene. This is the first successful attempt to synthesize the trimannose core in bacteria, and as such, this study opened several interesting possibilities, like further elongation of this basic structure to obtain more complex human-like glycans.

To reach this goal, expression of additional eukaryotic glycosyltransferases is required, and once again, *S. cerevisiae* is an ideal source for the required enzymes. Summarizing, synthesis of *N*-linked glycans in *S. cerevisiae* starts on the cytoplasmic face of the endoplasmic reticulum (ER) where several glycosyltransferases (Alg7, Alg13, Alg14, Alg1, Alg2 and Alg11) synthesize a glycan comprised of 2 *N*-acetylglucosamine (GlcNAc) residues and 5 mannose (Man) residues [146]. It is important to address here that this $Man_5GlcNAc_2$ structure is just 2 Man residues away from the trimannose core ($Man_3GlcNAc_2$), and those 2 Man residues are transferred by a single α1,2-mannosyltransferase [122], Alg11, making it relatively simple to extend the current trimannose synthesis process for production of $Man_5GlcNAc_2$. Figure 3.1 shows the hypothetical pathway for synthesis of $Man_5GlcNAc_2$ in *E. coli*.

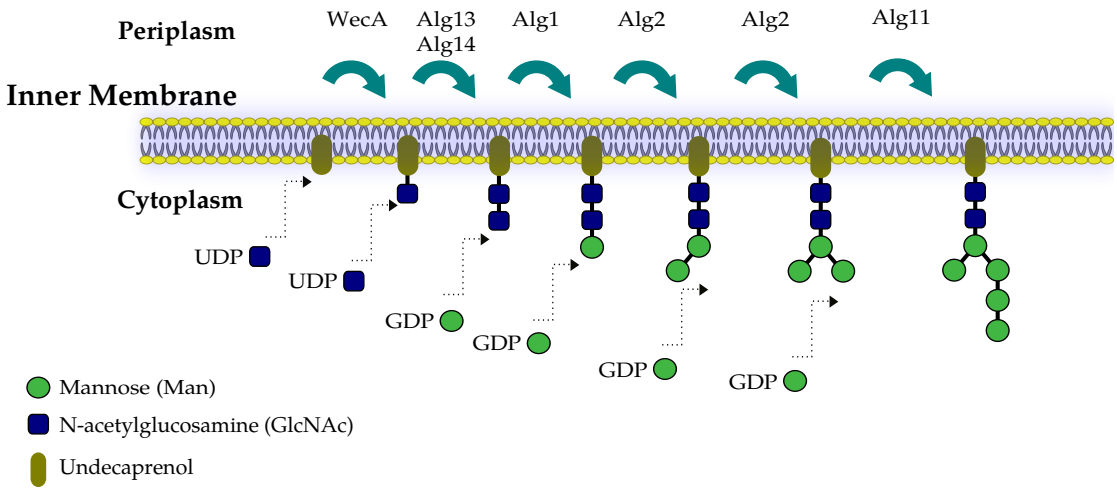

Figure 3.1: *Hypothetical pathway for synthesis of $Man_5GlcNAc_2$ in E. coli.* This picture shows a schematic for glycan synthesis in *E. coli* inner membrane. Previously, expression of Alg13, Alg14, Alg1 and Alg2 from *S. cerevisiae* in *E. coli* permitted the synthesis of $Man_3GlcNAc_2$. Now we hypothesize that adding Alg11 to the system will result on the synthesis of $Man_5GlcNAc_2$.

Alg11 is a membrane protein. Previous studies using fusion analysis suggest that it has an *N*-terminal transmembrane domain, 2 membrane associated domains and a *C*-terminal soluble domain [1]. Interestingly, it has been previously expressed in *E. coli* as a Thioredoxin *C*-terminal fusion and proven to be active *in vitro* when extracted as part of the membrane fraction [122]. However, no evidence of *in vivo* activity in *E. coli* has been reported before this study.

It is also worth mentioning that terminal α1,2-linked mannose is a recurring feature in some human pathogens [129, 132] and the presence of this structure on the HIV virus envelope is of particular importance, given several anti-HIV drugs, currently on testing, bind to terminal α1,2-linked mannose residues [11, 12, 74, 189]. Taking this into account, the product of Alg11 becomes a particularly interesting glycan because of its potential use in vaccine research and development.

3.2 Materials and methods

3.2.1 Bacterial strains and media

Antibiotic selection was maintained at: 25 μg/mL chloramphenicol (Cam) and 50 μg/mL kanamycin (Kan). Luria Bertani (LB) media was used for *E. coli*, supplemented with glucose at 0.2% as indicated. Protein expression was induced by adding L-arabinose and isopropyl -d-thiogalactoside (IPTG) at 0.2% and 100 mM, respectively. Yeast FY834 was maintained on YPD media and synthetic-defined-Uracil media was used to select or maintain yeast plasmids. E. coli MC4100 (F- araD139 Δ(argF-lac)U169 flbB5301 deoC1 ptsF25 relA1 rbsR22 rpsL150 thiA) was used as the recipient strain for genetic manipulations. To delete the *gmd* gene in MC4100 cells, P1*vir* phage transduction was performed using standard methods and the Keio collection [10].

3.2.2 Plasmid construction

Plasmids were constructed using standard homologous recombination in *S. cerevisiae* as previously described [149]. Briefly, pMWO7-YCG (see Fig. 1) was constructed by first PCR amplifying the genes *alg13*, *alg14*, *alg1*, *alg2*, and *alg11* from *S. cerevisiae* genomic DNA with primers containing appropriate regions of overlap. The gene encoding for the Mistic protein was PCR amplified in a similar way from a plasmid template. PCR products and the linearized vector pMWO7 (p15a origin and Cam resistance) were used to transform yeast strain FY834. Constructs assembled in yeast were electroporated into E. coli for verification via PCR, restriction enzyme digestion and/or sequencing. pBAD(*trxA:alg11:his*$_6$)-DEST49 was a generous gift from Dr. Barbara Imperiali.

Diagrams of the plasmids used in this study are shown in figure 3.2.

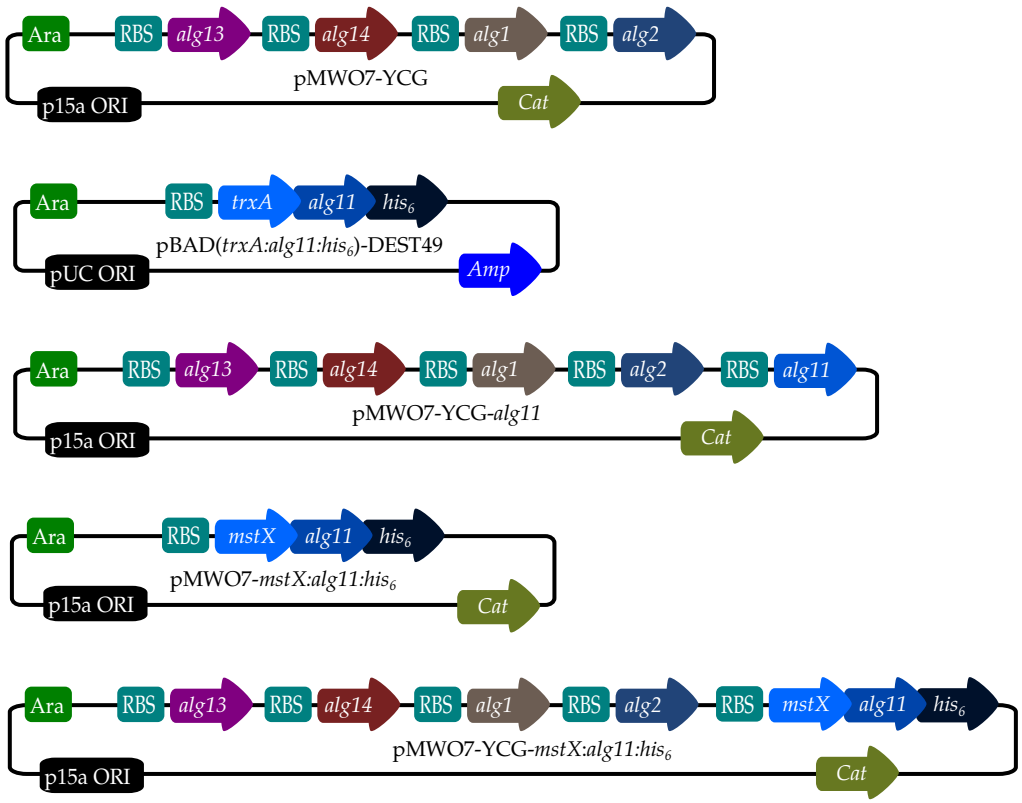

Figure 3.2: *Plasmids used in this study.*

3.2.3 Isolation of membranes and LLOs

Membranes were isolated similarly to the method of Marani et al [106]. For LLOs extraction, overnight cultures were diluted 1:100 into LB broth and incubated at 37°C with shaking (250 rpm) until reaching A600 ~0.6. Protein expression was then induced with 0.2% *L*-arabinose and cells were incubated at 30°C overnight in screw-capped flasks filled with media. Lipid-linked oligosaccharides (LLOs) were extracted and partially purified as described previously [56]. Briefly, the LLO extraction protocolwas based on a total lipid extraction protocol using chloroform:water (2:1) as described previously [53]. This solvent mix-

ture extracts the most hydrophobic lipids, leaving LLOs in the precipitate. The precipitate was washed with water to remove hydrophilic molecules and then LLOs were extracted preferentially using chloroform:methanol:water (10:10:3). Contaminating oligosaccharides not attached to lipids were removed by adsorbing extracted LLOs on DEAE-cellulose. LLOs were then recovered from DEAE-cellulose using an ammonium acetate solution. After releasing glycans from LLOs by acid hydrolysis, a butanol/water mixture was used to extract glycans to the water phase while lipids remained in the butanol phase. Finally, two ionic exchange resins (AG50W-X8 (Sigma) and AG1-X8 (BioRad)) were used for removing the remaining salts.

3.2.4 Western blot analysis

Purified protein samples were separated using 12% SDS-PAGE gels and transferred to PVDF membranes. Proteins that harbored 6x-His affinity tags were detected with a monoclonal anti-His (C-term) antibody (Invitrogen), per manufacturers instructions.

3.2.5 Glycans characterization by mass spectrometry (MS)

Glycans were released from E. coli LLOs as described previously [56]. Some glycans were further digested with *C. ensiformis* α-exomannosidase in 50 mM ammonium acetate, pH 4.5, followed by further desalting steps using a carbon column (envi-carb SPE tube, Supelco). Permethylation of glycans was performed as described [79]. MALDI-MS profiles were acquired using a 4800 MALDI-TOF/TOF (Applied Biosystems) in positive ion reflectron mode (500-4000 Da mass range). MALDI-MS/MS sequencing of permethylated glycans was per-

formed in positive ion reflectron mode. Air was used as the collision gas. Structure assignment was according to the method of Yu et al [188].

3.2.6 Fluorophore assisted carbohydrate electrophoresis (FACE)

Glycans were derivatized with 5μL 7amino1,3naphthalenedisulfonic acid (ANDS) 0.15M in 15% acetic acid and 5μL of 1M sodium cyanoborohydride in DMSO, and electrophoresed through a 14% acrylamide gel using 0.01% thorin in 20% glycerol as loading dye, and 1.92M glycine in 0.25M Tris base (pH 8.3) as running buffer at 10mA constant current [56]. Derivatized glycans were detected as fluorescent bands using a UV transilluminator.

3.3 Results

3.3.1 Expression of Alg11 in *E. coli*

Expression of Alg11 was attempted using 3 different strategies: from the wild type sequence (from *S. cerevisiae*), as a *C*-terminal fusion to Thioredoxin (TrxA), and as *C*-terminal fusion to Mistic (MstX). As reported by O'Reilly et al [122], when using the wild type sequence for Alg11, western blot analysis shows no detectable amounts of this protein. However, when expressed as a *C*-terminal fusion to TrxA, Alg11 can be detected (83kDa) and it localizes to the membrane fraction of MC4100 *gmd::kan* cells (figure 3.3a). In the MstX:Alg11 case, the fusion is also expressed and can be detected in the membrane fraction (77kDa). In both cases, there seem to be at least 4 lower molecular weight bands presumably corresponding to degradation products.

3.3.2 *In vivo* activity of Alg11

A previous study showed that $Man_3GlcNAc_2$ can be synthesized in *E.coli* given an appropriate set of glycosyltransferases is expressed in MC4100 *gmd::kan* [167]. Based on those results, *in vivo* activity of Alg11 in *E. coli* can be assessed taking advantage of the fact that $Man_3GlcNAc_2$ is the natural substrate for Alg11 and that MC4100 *gmd::kan* accumulates GDP-Man[1] in the cytoplasm thanks to the *gmd* deletion. In this case, the use of lectins like ConA (from *Canavalia ensiformis*) for glycan analysis is not the best option, taking into account that ConA reacts in a similar manner with $Man_3GlcNAc_2$ and $Man_5GlcNAc_2$, so it cannot be used to discern between these two structures.

Given the above, primary glycan analysis was performed using FACE followed by MS for confirmation of results.

Glycan analysis

For evaluation of the *in vivo* activity of TrxA:Alg11 we used plasmid pMWO7-YCG for synthesis of the trimannose core and expressed TrxA:Alg11 from plasmid pBAD(*trxA:alg11:his*$_6$)-DEST49. For Alg11 and MstX:Alg11 evaluation, plasmids pMWO7-YCG-*alg11* and pMWO7-YCG-*mstX:alg11* were constructed (figure 3.2), so all glycosyltransferases were expressed from a single plasmid.

Unexpectedly, in opposition to previous experiments *in vitro* [122], TrxA:Alg11 did not show any detectable activity *in vivo*. Given TrxA is a soluble protein, it might be interfering with proper insertion into the membrane, although TrxA:Alg11 hydrophobic regions could still interact non-specifically with the membrane, explaining why TrxA:Alg11 shows up in the membrane fraction.

[1] This is the sugar-nucleotide donor for Alg11

On the other hand, wild type Alg11 shows some *in vivo* activity, nevertheless it is not 100% efficient as FACE analysis suggests (figure 3.3b). Further MS analysis showed that Alg11 efficiency could be around 15% when comparing the peak corresponding to $Hex_5HexNAc_2$ with the $Hex_3HexNAc_2$ peak (figure 3.4a), which is in some way an unexpected result given Alg11 expression escapes immunoblotting detection. Nevertheless, this would explain why efficiency is still on the low side and, since glycan population is too much heterogeneous for some practical purposes, improvement of the system is desirable.

Taking this into account, and in an effort to integrate Alg11 properly to the membrane, our next step was to try *C*-terminal fusions to membrane integration partners such as GlpF and MstX. While GlpF fusion did not express very well (data not shown), MstX:Alg11 showed good expression and was fully active in MC4100 *gmd::kan* (figure 3.3a and b). Surprisingly, in this case the prevalent glycoform was $Man_5GlcNAc_2$. Furthermore, as MS analysis confirms (figure 3.4b), all precursors to $Man_3GlcNAc_2$ appeared at nearly undetectable amounts, resulting in not just a high MstX:Alg11 efficiency but also a nearly 100% efficiency for the whole pathway.

Mannosidase analysis

Given the inherent limitations in MS and FACE analysis respect to identifying linkages between sugar residues, isolated glycans were digested using a α1,2-mannosidase. As Alg11 attaches 2 mannose residues with specific α1,2 linkages, digestion of $Man_5GlcNAc_2$ with α1,2-mannosidase should result on production of $Man_3GlcNAc_2$. FACE analysis of digested and undigested glycans produced on cells expressing MstX:Alg11[2] resulted on a banding pattern con-

[2]This experiment was performed by our colaborators at Glycobia Inc.

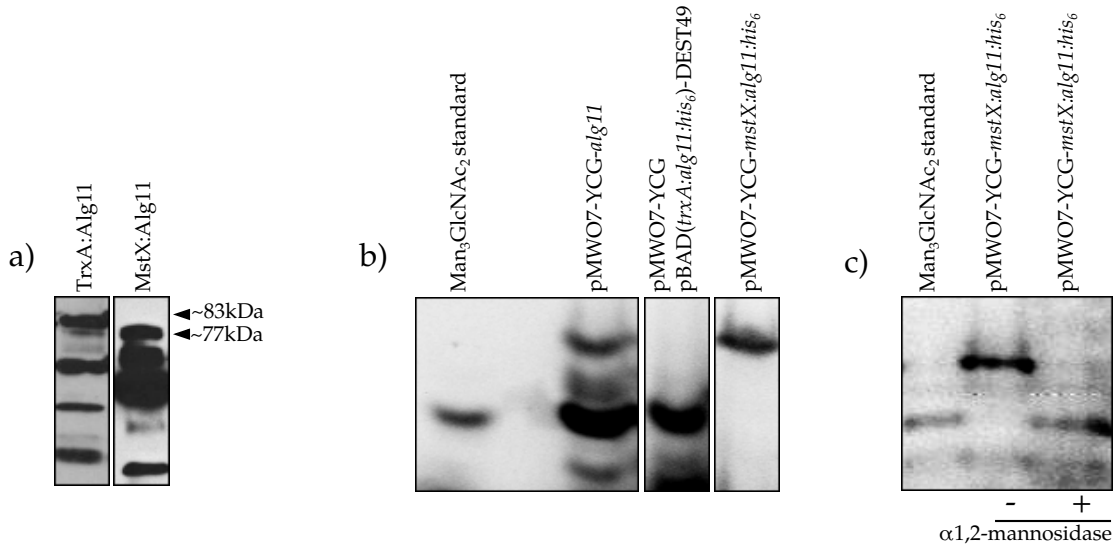

Figure 3.3: *Alg11 expression and in vivo activity assessment in E. coli*. (a) Expression of Alg11 fusions in MC4100 *gmd::kan*. TrxA:Alg11 was expressed from pBAD(*trxA:alg11:his*$_6$)-DEST49, while MstX:Alg11 was expressed from pMWO7-*mstX:alg11*. (b) FACE analysis of several Alg11 constructs activity *in vivo*. TrxA:Alg11 did not show any detectable activity while MstX:Alg11 showed the highest efficiency. (c) α1,2-mannosidase digestion of *in vivo* produced glycans after MstX:Alg11 expression. FACE results are consistent with removal of 2 mannose residues from $Man_5GlcNAc_2$ to produce $Man_3GlcNAc_2$.

sistent with specific degradation of $Man_5GlcNAc_2$ to $Man_3GlcNAc_2$, suggesting that the product of MstX:Alg11 catalytic activity is indeed $Man_5GlcNAc_2$ (figure 3.3c).

3.4 Discussion

Alg11 appears not to have any signal peptide compatible with membrane insertion in gram negative bacteria[3] , so its insertion into the membrane might not obey to interactions with the translocation machinery but might be due to hydrophobic interactions related to peripheral membrane associated domains that are not necessarily spanning the membrane. This hypothesis could explain why

[3]Sequence was analized using SignalP 4.0 server [130]

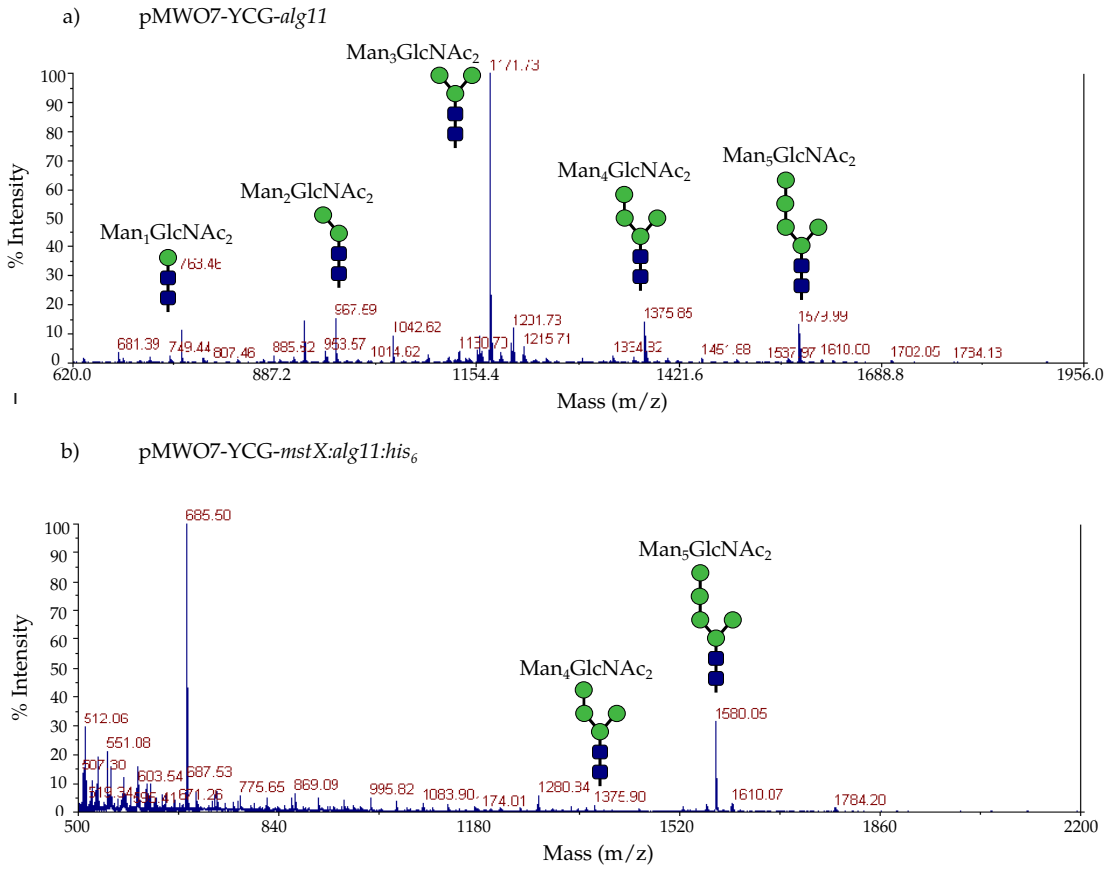

Figure 3.4: *MS analysis of glycans extracted from E. coli cells expressing glycosyltransferases for synthesis of $Man_5GlcNAc_2$.* Analysis of glycans extracted from MC4100 *gmd::kan* cells expressing glycosyltransferases for synthesis of $Man_3GlcNAc_2$ and (a) wild type Alg11, or (b) MstX:Alg11. These results confirm that Alg11 is active in *E. coli* as well as show that *N*-terminal MstX fusion enhance not only Alg11 expression but also the overall pathway performance.

Alg11 manages to attain an active conformation instead of aggregating in the cytoplasm. Also, since TrxA is a soluble protein and does not have any translocation signal sequence, this Alg11 peripheral membrane associated domains could be responsible for localization of TrxA:Alg11 to the membrane fraction, hence interaction with the translocation machinery would not be required.

Surprisingly, TrxA:Alg11 do not show any detectable activity *in vivo*. Given previous *in vitro* experiments based on *E. coli* expressed TrxA:Alg11 showed the

expected catalytic activity [122], failure to attach additional mannose residues to $Man_3GlcNAc_2$ *in vivo* could be a consequence of improper folding or relatively poor expression. It is possible that during the *in vitro* experiments an excess of protein was used, compared to its *in vivo* availability in *E. coli*. Additionally, the *in vitro* reaction mixture included NP-40 (nonyl phenoxypolyethoxylethanol) which might play a role on stabilizing the TrxA:Alg11 structure and spatial conformation.

It is also worth noticing that Alg11, even though its expression cannot be detected by immunoblotting, shows catalytic activity *in vivo*. This suggest that at least a small fraction of Alg11 reaches a functional configuration, being this small amount enough to detect activity in MC4100 *gmd::kan* cells. Nevertheless, $Man_3GlcNAc_2$ elongation efficiency is relatively low, possibly because of poor expression of this glycosyltransferase. Given that, as a stabilization and membrane insertion strategy, *N*-terminal fusion to MstX was attempted. The expression profile showed that cellular amounts of MstX:Alg11 were comparable to TrxA:Alg11. Also, membrane localization displayed a similar pattern to TrxA:Alg11, where MstX:Alg11 was successfully integrated to the membrane.

Fusion to MstX was expected to enhance membrane association and serve as folding partner to Alg11. From western blot analysis, it is hard to establish any impact MstX might be having on these two parameters inasmuch as there are not any evident differences between TrxA:Alg11 and MstX:Alg11 banding patterns. However, from FACE and MS analysis, it becomes evident that MstX fusion is having a positive effect on either Alg11 folding or membrane integration, or maybe both. Additionally, results indicated that synthesis of $Man_5GlcNAc_2$ has a near 100% efficiency with $Man_3GlcNAc_2$ and its precursors being at almost undetectable levels, suggesting that MstX fusion is not only beneficial for

Alg11 expression and activity but somehow it is also affecting the performance of the whole pathway.

Summarizing, the above results showed that fusion to Mstx can enhance Alg11 expression and activity. Nevertheless, it is difficult to decouple the membrane insertion from the folding enhancement effects given MstX could be impacting both at the esame time. On the other hand, the unexpected lack of accumulation of $Man_3GlcNAc_2$ and its precursors could be hypothetically related to the native interactions Alg1-Alg2 and Alg1-Alg11 observed in *Saccharomyces cerevisiae* [58]. As a result of these hypothetical interactions, an Alg1-Alg2-MstX:Alg11 complex could be having a positive kinetic effect, making the whole glycan synthesis process more efficient. If so, MstX could be responsible for enhanced membrane insertion, hence facilitating the required interactions for assembly of this molecular machine.

CHAPTER 4

FUNCTIONAL EXPRESSION OF GNTI, A HUMAN GLYCOSYLTRANSFERASE, IN *ESCHERICHIA COLI*

4.1 Introduction

Recent advances in glycobiology have demonstrated the importance of *N*-linked glycosylation for therapeutic proteins [154], being *N*-glycans structure and composition of high relevance respect to protein folding and functionality, as well as playing a role on targeting this proteins to specific tissues [15, 154, 163, 151]. With this in mind, industrial scale production of therapeutic glycoproteins is of high interest but it is also challenge, mostly because synthesis of glycoproteins suitable for human use is restricted to human cell lines and, in some cases, to modified mammalian cell lines [45, 126]. In general, while some of these cell lines show an adequate performance respect to production, there are several drawbacks inherent to mammalian cells like the need for expensive media, risk of viral contamination, and relatively difficult genetic manipulation [38]. Because of that, alternative systems for therapeutic glycoprotein production are being investigated.

Some examples of these emerging systems are genetically modified plant cells, filamentous fungi, insect cells and yeast cells [45]. Nevertheless, an obstacle for the implementation of these alternative systems resides in the difficulty associated with the synthesis of human-like *N*-glycans that are often necessary to achieve proper functionality and maintain adequate serum half-life of therapeutic glycoprotein [163]. While most eukaryotic *N*-glycans, including human glycans, share a common core structure known as the trimannose core ($Man_3GlcNAc_2$), their structures diverge widely and most of the time these

non-human *N*-glycans result in immunogenicity of their glycoprotein carriers [126, 135, 162] or poor stability in the blood torrent [15]. Because of this problem, to produce therapeutic glycoproteins in non-human cells, synthesis of human-like *N*-glycans is of critical importance, resulting in a number of studies focused on this goal using diverse types of eukaryotic organisms [26, 69, 78, 180]. Notably, in most of these studies recombinant expression of human glycosyltransferases has been the strategy of choice.

Given glycosyltransferases are membrane proteins, their recombinant expression is often difficult [16, 43, 50, 100, 113, 175] and usually they require engineering of the host as well as engineering of the protein itself. Particularly, for membrane protein engineering, strategies including fusion to folding partners and membrane integration partners, like MBP and Mistic, have proven to be very useful [72, 113, 120, 139, 184]. Still, selection of the best engineering strategy is not trivial and often rational design is not enough for ensuring proper expression, mainly given our limited understanding of protein folding and the membrane integration process.

Despite those difficulties, recombinant expression of glycosyltransferases has been achieved in several cases, being of particular relevance for this study the expression of several glycosyltransferases from *S. cerevisiae* for synthesis of the eukaryotic trimannose core in *E. coli* [167]. In summary, $Man_3GlcNAc_2$ was synthesized by expression of Alg13, Alg14, Alg1 and Alg2 alongside with deletion of the *gmd* gene from the bacterial host. Interestingly, this development opened the possibility for synthesizing complex human-like glycans in *E. coli* and hence it constitutes a new potential alternative for the production of therapeutic glycoproteins at high scale.

Among the most relevant *N*-glycans required for therapeutic applications

we find terminally sialylated glycans. In human cells, synthesis of sialylated biantennary glycans happens in the Golgi apparatus. They are derived from highly mannosylated glycans which, after being attached to proteins in the endoplasmic reticulum, are trimmed down to a $Man_5GlcNAc_2$ structure by several α-mannosidases [78]. Thereafter, *N*-acetylglucosaminyltransferase I (GnTI) transfers a GlcNAc residue on top of this structure [54] and then 2 additional mannose residues are removed. Finally, a biantennary structure is built having $Man_3GlcNAc_2$ at the core [70]. Interestingly, GnTI can also catalyze the addition of a GlcNAc unit to $Man_3GlcNAc_2$ instead of $Man_5GlcNAc_2$, although with a reduced efficiency [54]. Based on this result, the possibility exists for $GlcNAcMan_3GlcNAc_2$ to be synthesized *in vivo* in *E. coli*, starting from $Man_3GlcNAc_2$ and then adding a GlcNAc residue by recombinant expression of GnTI in the cytoplasm. A scheme for this hypothetical pathway in *E. coli* is shown in figure 4.1

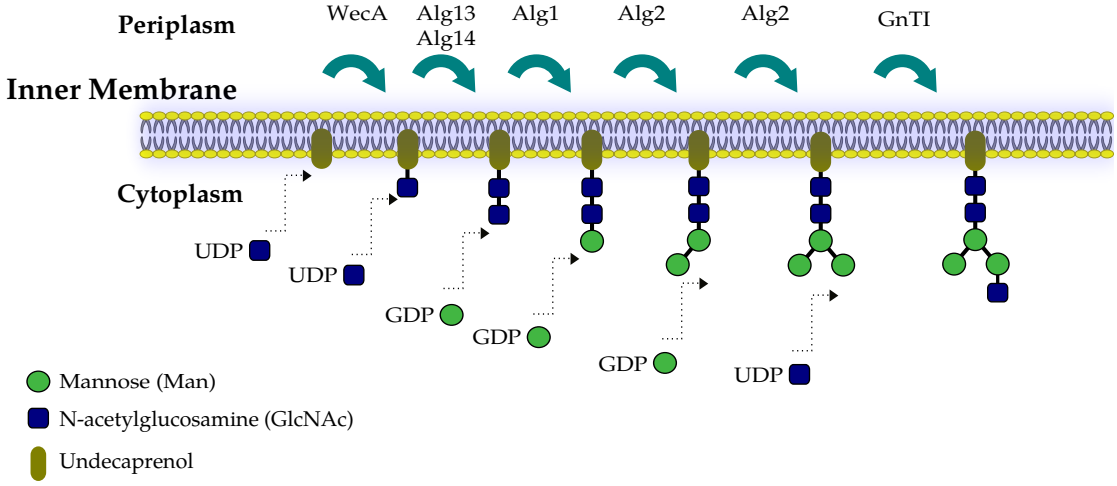

Figure 4.1: *Hypothetical pathway for synthesis of $GlcNAcMan_3GlcNAc_2$ in E. coli.* This picture shows a schematic for glycan synthesis in *E. coli* inner membrane. Previously, expression of Alg13, Alg14, Alg1 and Alg2 from *S. cerevisiae* in *E. coli* permitted the synthesis of $Man_3GlcNAc_2$. Now we hypothesize that adding GnTI to the system will result on the synthesis of $GlcNAcMan_3GlcNAc_2$.

Notably, human GnTI has been previously expressed in the cytoplasm of *E. coli* as a *C*-terminal fusion to MBP, and its activity demonstrated *in vitro* after purification and immobilization [54]. It is also interesting that 106 amino acids from the *N*-terminus can be removed while maintaining catalytic activity. Given GnTI is a type I membrane protein, these 106 amino acids are probably corresponding to the stem and transmembrane domain, which suggests that GnTI catalytic site resides in the *C*-terminal soluble domain [143].

In this study we explored different alternatives for functional expression of GnTI in *E. coli* and, for the first time, we were able to demonstrate *in vivo* activity of this enzyme by expression of glycosyltransferases from *S. cerevisiae* and use of the resulting $Man_3GlcNAc_2$ structure as substrate for the synthesis of $GlcNAcMan_3GlcNAc_2$.

4.2 Materials and methods

4.2.1 Bacterial strains and media

Antibiotic selection was maintained at: 25 μg/mL chloramphenicol (Cam) and 50 μg/mL kanamycin (Kan). Luria Bertani (LB) media was used for *E. coli*, supplemented with glucose at 0.2% as indicated. Protein expression was induced by adding L-arabinose and isopropyl -d-thiogalactoside (IPTG) at 0.2% and 100 mM, respectively. Yeast FY834 was maintained on YPD media and synthetic-defined-Uracil media was used to select or maintain yeast plasmids. E. coli MC4100 (F- araD139 Δ(argF-lac)U169 flbB5301 deoC1 ptsF25 relA1 rbsR22 rpsL150 thiA) was used as the recipient strain for genetic manipulations. To delete the *gmd* gene in MC4100 cells, Shuffle® T7, Shuffle® T7 express, Origami®

2, and DR473, P1*vir* phage transduction was performed using standard methods and the Keio collection [10].

4.2.2 Plasmid construction

Plasmids were constructed using standard homologous recombination in *S. cerevisiae* as previously described [149]. Briefly, pMWO7-YCG (see Fig. 1) was constructed by first PCR amplifying the genes *alg13*, *alg14*, *alg1*, and *alg2* from *S. cerevisiae* genomic DNA with primers containing appropriate regions of overlap. The genes encoding for Mistic and GnTI were PCR amplified in a similar way from a plasmid template (ATCC MGC-2304). PCR products and the linearized vector pMWO7 (p15a origin and Cam resistance) were used to transform yeast strain FY834. Constructs assembled in yeast were electroporated into E. coli for verification via PCR, restriction enzyme digestion and/or sequencing. Diagrams of the plasmids used in this study are shown in figure 4.2.

4.2.3 Isolation of membranes and LLOs

Membranes were isolated similarly to the method of Marani et al [106]. For LLOs extraction, overnight cultures were diluted 1:100 into LB broth and incubated at 37°C with shaking (250 rpm) until reaching A600 0.6. Protein expression was then induced with 0.2% *L*-arabinose and cells were incubated at 30°C overnight in screw-capped flasks filled with media. Lipid-linked oligosaccharides (LLOs) were extracted and partially purified as described previously [56]. Briefly, the LLO extraction protocolwas based on a total lipid extraction protocol using chloroform:water (2:1) as described previously [53]. This solvent mixture extracts the most hydrophobic lipids, leaving LLOs in the precipitate. The

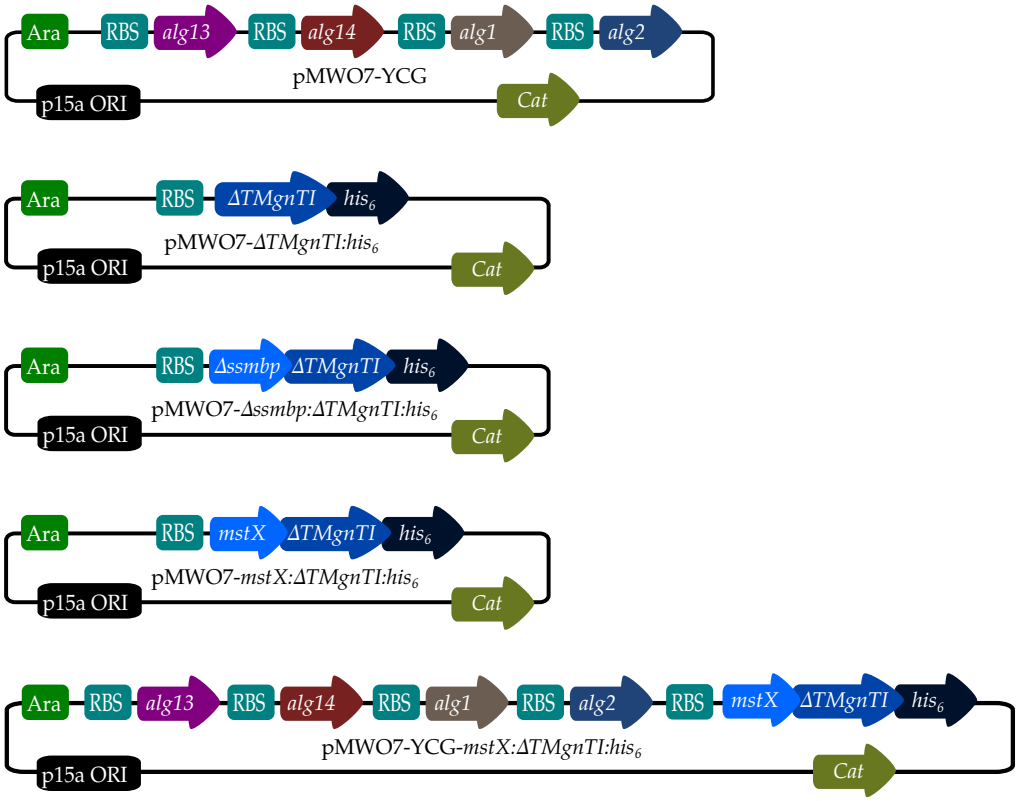

Figure 4.2: *Plasmids used in this study.*

precipitate was washed with water to remove hydrophilic molecules and then LLOs were extracted preferentially using chloroform:methanol:water (10:10:3). Contaminating oligosaccharides not attached to lipids were removed by adsorbing extracted LLOs on DEAE-cellulose. LLOs were then recovered from DEAE-cellulose using an ammonium acetate solution. After releasing glycans from LLOs by acid hydrolysis, a butanol/water mixture was used to extract glycans to the water phase while lipids remained in the butanol phase. Finally, two ionic exchange resins (AG50W-X8 (Sigma) and AG1-X8 (BioRad)) were used for removing the remaining salts.

4.2.4 Western blot analysis

Purified protein samples were separated using 12% SDS-PAGE gels and transferred to PVDF membranes. Proteins that harbored 6x-His affinity tags were detected with a monoclonal anti-His (C-term) antibody (Invitrogen), per manufacturers instructions.

4.2.5 Glycans characterization by mass spectrometry (MS)

Glycans were released from E. coli LLOs as described previously [56]. Some glycans were further digested with *C. ensiformis* α-exomannosidase in 50 mM ammonium acetate, pH 4.5, followed by further desalting steps using a carbon column (envi-carb SPE tube, Supelco). Permethylation of glycans was performed as described [79]. MALDI-MS profiles were acquired using a 4800 MALDI-TOF/TOF (Applied Biosystems) in positive ion reflectron mode (500-4000 Da mass range). MALDI-MS/MS sequencing of permethylated glycans was performed in positive ion reflectron mode. Air was used as the collision gas. Structure assignment was according to the method of Yu et al [188].

4.2.6 Fluorophore assisted carbohydrate electrophoresis (FACE)

Glycans were derivatized with 5μL 7amino1,3naphthalenedisulfonic acid (ANDS) 0.15M in 15% acetic acid and 5μL of 1M sodium cyanoborohydride in DMSO, and electrophoresed through a 14% acrylamide gel using 0.01% thorin in 20% glycerol as loading dye, and 1.92M glycine in 0.25M Tris base (pH 8.3) as running buffer at 10mA constant current [56]. Derivatized glycans were detected as fluorescent bands using a UV transilluminator.

4.3 Results

4.3.1 Expression of GnTI in *E. coli*

Three different approaches were evaluated for human GnTI (hGntI) expression in MC4100 *gmd::kan*. Since expression of full lenght hGnTI has previously proven to be difficult [54], probably because of its transmembrane domain and the fact that it does not have any signal sequence for membrane integration in gram negative bacteria[1], our starting point was a truncated hGnTI (ΔTMGnTI) lacking the first 106 *N*-terminal amino acids [143]. Plasmid pMWO7-*ΔTMgnTI:his*$_6$ was used for assessing ΔTMGnTI expression. As shown in figure 4.3a, ΔTMGnTI (43kDa) expression is poor and several degradation bands were detected, suggesting that even without its trasnmembrane domain and stem ΔTMGnTI is highly unstable in MC4100 *gmd::kan*.

Following the strategy presented in [54], we proceeded to fuse ΔTMGnTI to the *C*-terminus of ΔssMBP to enhance its solubility and stability. Plasmid pMWO7-*Δssmbp:ΔTMgnTI:his*$_6$ was used for expression of ΔssMBP:ΔTMGnTI (88kDa). This fusion was successfully detected when expressed in MC4100 *gmd::kan*, although considerable degradation was detected as well (figure 4.3a).

A third approach, contrasting with the ΔssMBP:ΔTMGnTI alternative, was to fuse ΔTMGnTI to the *C*-terminus of MstX. In this case, MstX serves not just as folding partner, but also anchors ΔTMGnTI to the membrane. Given glycan synthesis happens on a lipid carrier embedded in the inner membrane, we hypothesized that MstX:ΔTMGnTI attachment to membrane should be kinetically advantageous. As shown in figure 4.3a, when expressing MstX:ΔTMGnTI (56kDa) from pMWO7-*mstX:ΔTMgnTI:his*$_6$, it can be detected in the mem-

[1]Sequence was analized using SignalP 4.0 server [130]

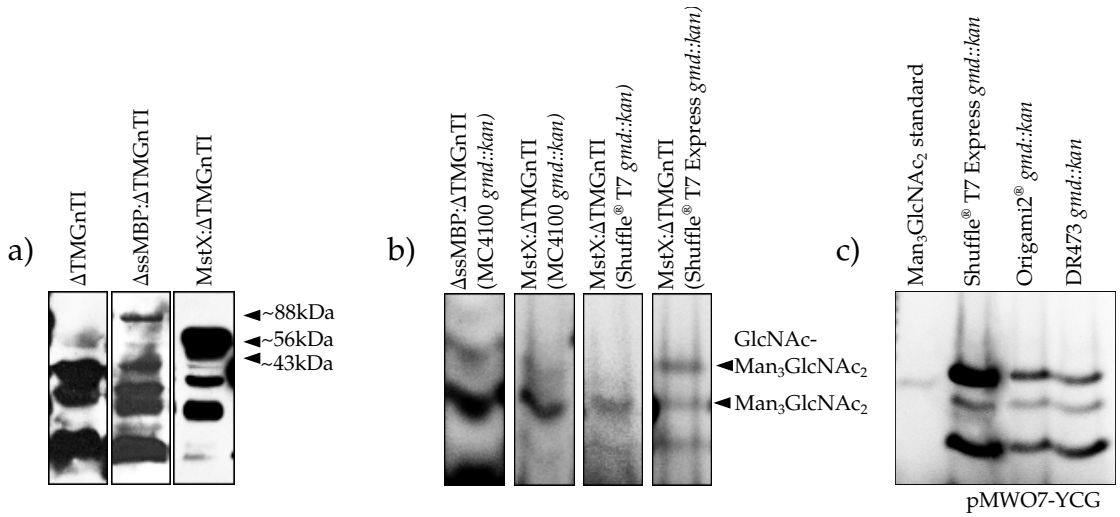

Figure 4.3: *hGnTI expression and in vivo activity assessment in E. coli.* (a) Expression of hGnTI in MC4100 *gmd::kan*. ΔTMGnTI was expressed from pMWO7-*ΔTMgnTI:his*$_6$ and ΔssMBP:ΔTMGnTI was expressed from pMWO7-*Δssmbp:ΔTMgnTI:his*$_6$. In both cases, whole cell lysate was used for immunoblotting. MstX:ΔTMGnTI was expressed from pMWO7-*mstX:ΔTMgnTI:his*$_6$ and the membrane fraction was used for immunoblotting. The evidence suggest that MstX fusion has a stabilizing effect on hGnTI. (b) FACE analysis of glycans extracted from several *E. coli* hosts expressing different hGnTI fusions. c) FACE analysis of glycans extracted from oxidizing cytoplasm *E. coli* strains expressing glycosyltransferases for synthesis of $Man_3GlcNAc_2$.

brane fraction. Notably, full lenght MstX:ΔTMGnTI appears as the prevalent band and, while degradation is still present, it is not as extended as in the ΔssMBP:ΔTMGnTI case.

4.3.2 *In vivo* activity of GnTI

Since ΔTMGnTI is poorly expressed in MC4100 *gmd::kan*, activity was only assessed for ΔssMBP:ΔTMGnTI and MstX:ΔTMGnTI. LLOs were extracted from 1L low oxygen cultures 22h after induction and analyzed using FACE and MS.

FACE analysis (figure 4.3b) indicates that ΔssMBP:ΔTMGnTI is active in MC4100 *gmd::kan*, although its efficiency is relatively low when comparing the

putative $GlcNAcMan_3GlcNAc_2$ band intensity to the $Man_3GlcNAc_2$ band. In a similar way, MstX:ΔTMGnTI did not show a high efficiency either and it appears to be even lower than ΔssMBP:ΔTMGnTI, despite the apparent advantage of MstX:ΔTMGnTI in regards to expression and lower degradation. MS analysis confirmed the presence of $HexNAcHex_3HexNAc_2$ (peak at 1416 in figure 4.4) in both ΔssMBP:ΔTMGnTI and MstX:ΔTMGnTI samples. However, it also shows a peak at 1375 corresponding to $Man_4GlcNAc_2$, probably a product from unspecific Alg1 or Alg2 activity [167]. Given this fact, it is also important to make clear that $Man_4GlcNAc_2$ can be detected using FACE, but given the proximity of the $Man_4GlcNAc_2$ to the $GlcNAcMan_3GlcNAc_2$ band, FACE analysis is not sufficient for demonstrating of $GlcNAcMan_3GlcNAc_2$ synthesis, so MS confirmation is always necessary in this particular case.

MS analysis is also useful for estimating the relative efficiency by direct comparison of peak heights. In both cases (figure 4.4a and 4.4b) the $HexNAcMan_3HexNAc_2$ peak is approximately 10% of the $Man_3GlcNAc_2$ peak. Given that, we hypothesized that the poor activity of these enzymes *in vivo* could be related to the lack of disulfide bond formation machinery in *E. coli* cytoplasm, a reason why oxidizing cytoplasm *E. coli* strains were considered as host.

4.3.3 GnTI shows enhanced activity in oxidizing cytoplasm *E. coli*

GnTI amino acid sequence includes 5 cysteine residues available for disulfide bond formation. Even though there is not conclusive evidence regarding disulfide bond presence in GnTI structure, previous attempts to produce this enzyme

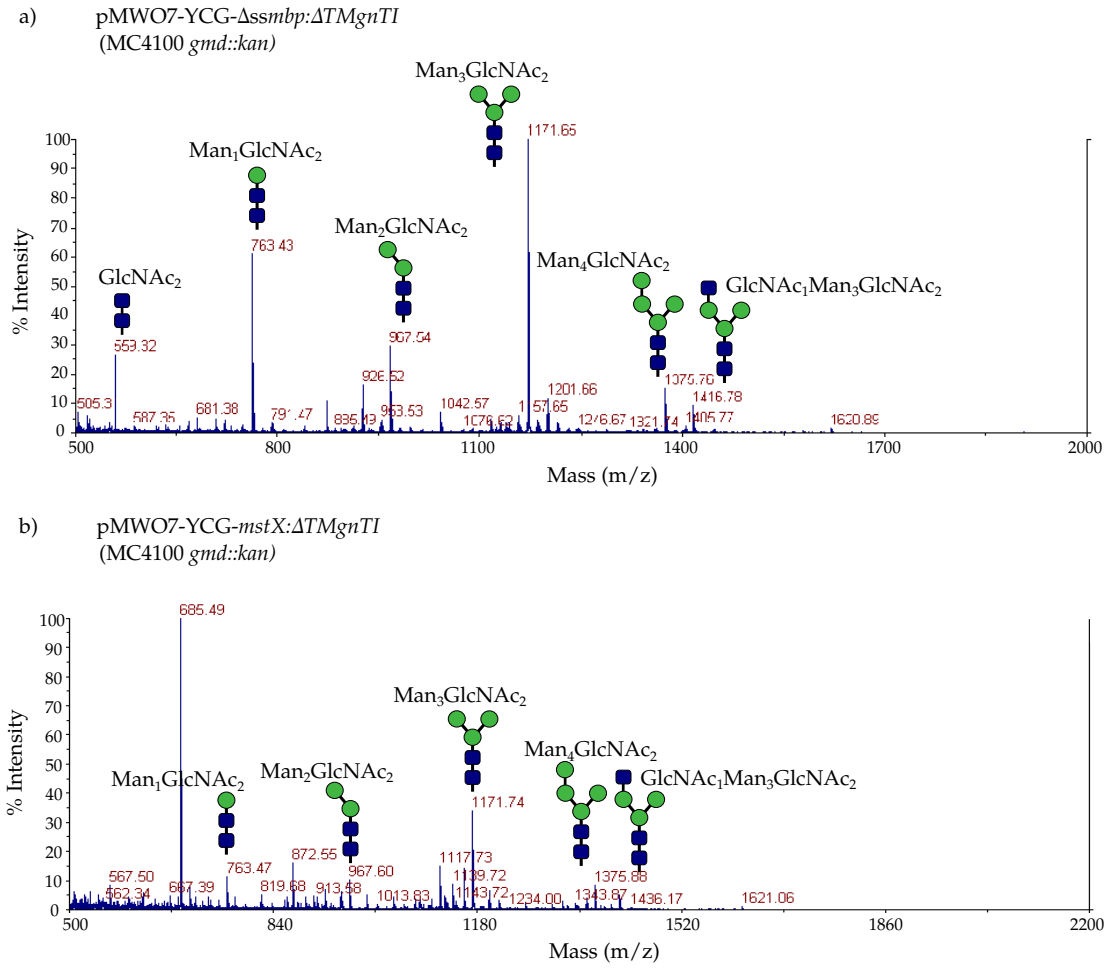

Figure 4.4: *MS analysis of glycans extracted from E. coli cells expressing glycosyltransferases for synthesis of $GlcNAcMan_3GlcNAc_2$.* (a) Glycans were extracted from MC4100 *gmd::kan* carrying pMWO7-YCG-*Δssmbp:ΔTMgnTI*. (b) Glycans were extracted from MC4100 *gmd::kan* carrying pMWO7-YCG-*mstX:ΔTMgnTI*. In both cases peaks consistent with GnTI enzymatic activity were found.

in *E. coli* showed that its solubility and activity benefits from expression in an oxidizing cytoplasm host [145].

Given the requirement for knocking out the *gmd* gene to enable MC4100 cells for synthesis of the trimannose core, we decided to delete this gene from some oxidizing cytoplasm *E. coli* strains: Shuffle® T7, Origami® 2 and DR473. Thereafter, plasmid pMWO7-YCG was used for evaluating production of the trimannose core in these modified strains using FACE (figure 4.3c). While trimannose

core synthesis was detected in all tested oxidizing cytoplasm strains, we limited GnTI evaluation to Shuffle® T7 *gmd::kan* and Shuffle® T7 express *gmd::kan*. For this experiment, plasmid pMWO7-YCG-*Δssmbp:ΔTMgnTI:his*$_6$ was used.

FACE analysis (figure 4.3b) suggests that MstX:ΔTMGnTI has little to no activity in Shuffle® T7 *gmd::kan*. However, its activity in Shuffle® T7 express *gmd::kan* appears to be better than in MC4100 *gmd::kan*. When comparing the upper band with the $Man_3GlcNAc_2$ band we can even infer that $GlcNAcMan_3GlcNAc_2$ production efficiency might be over 50%. Further MS analysis suggests that MstX:ΔTMGnTI is active in Shuffle® T7 *gmd::kan*, despite negative FACE results. Moreover, it appears to have approximately 50% efficiency (figure 4.5a). On the other hand, its activity in Shuffle® T7 express *gmd::kan* matches FACE results, confirming that $GlcNAcMan_3GlcNAc_2$ production efficiency surpasses 50% (figure 4.5b).

4.4 Discussion

As an eukaryotic membrane protein, hGnTI is difficult to express in *E. coli* given the lack of bacterial signal sequence for membrane translocation and the absence of specific eukaryotic chaperones in this host [174]. Fortunately, the membrane associated domain in GnTI is not required for enzymatic activity which permits engineering of GnTI as a soluble cytoplasmic protein in bacteria. Nevertheless, a version of GntI lacking the first 106 amino acids (ΔTMGnTI) failed to express in MC4100 *gmd::kan* cytoplasm and instead several bands possibly corresponding to degradation products were detected when immunoblotting. To overcome this problem, two engineering alternatives were tested. First, fusion to the soluble folding partner MBP (ΔssMBP:ΔTMGnTI) for expression of GnTI

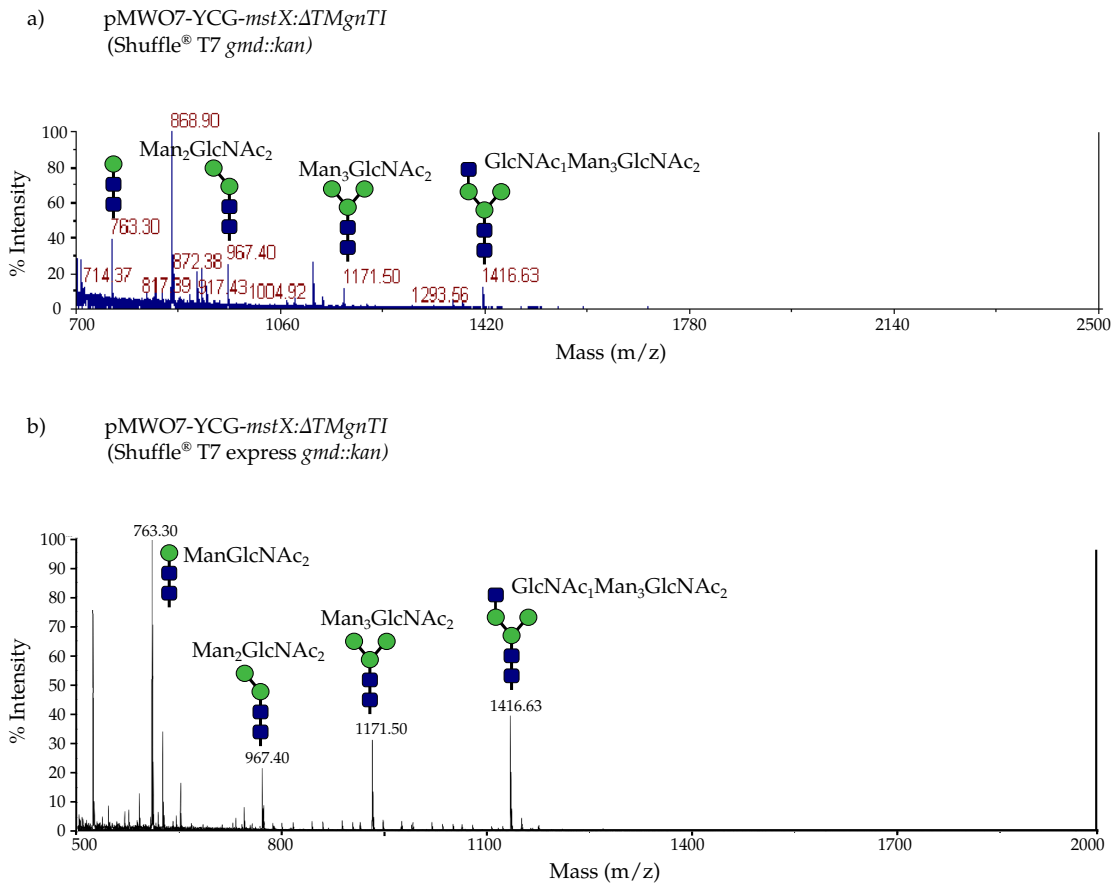

Figure 4.5: *MS analysis of glycans extracted from oxidizing cytoplasm E. coli cells expressing glycosyltransferases for synthesis of $GlcNAcMan_3GlcNAc_2$.* (a) Glycans were extracted from Shuffle® T7 *gmd::kan* carrying pMWO7-YCG-*Δssmbp:ΔTMgnTI*. (b) Glycans were extracted from Shuffle® T7 express *gmd::kan* carrying pMWO7-YCG-*mstX:ΔTMgnTI*. This analysis suggest that an oxidizing cytoplasm impacts positively hGnTI catalytic activity, possibly because it facilitates formation of disulfide bonds.

as a soluble cytoplasmic protein. Second, we rationalized that proper targeting of GnTI to *E. coli* inner membrane could be beneficial for reaction kinetics since glycan synthesis occurs over a lipid carrier embedded in the membrane. Given that, we decided that MstX (Mistic) had a great potential as folding partner and membrane anchor. Also, since its *C*-terminus faces the cytoplasm, we expected that the fusion protein (MstX:ΔTMGnTI) localized the enzymatic hGnTI subunit towards the cytoplasm.

Both fusion attempts resulted successful respect to protein expression, even though considerable degradation was detected. Still, MstX:ΔTMGnTI seemed to be a better option given immunoblotting analysis indicated that the full sized fusion protein was the prevalent band and degradation was not as extended, while ΔssMBP:ΔTMGnTI appeared less stable. MstX:ΔTMGnTI also localized to the membrane fraction as intended. Furthermore, for the first time, we were able to show *in vivo* activity of hGnTI in *E. coli* even though efficiency was relatively low, being ΔssMBP:ΔTMGnTI slightly better than MstX:ΔTMGnTI. While this is an interesting result from the scientific point of view, from the engineering perspective there is still space for improvement.

Given some evidence suggesting that hGnTI might have disulfide bonds in its structure [145] and the fact that it has a number of cysteine residues, based on previous observations [145], we decided to engineer oxidizing cytoplasm strains for production of eukaryotic-like *N*-glycans. Surprisingly $GlcNAcMan_3GlcNAc_2$ was not detected in Shuffle® T7 *gmd::kan* after FACE analysis. However, MS analysis suggested that MstX:ΔTMGnTI is active in this strain. On the other hand, Shuffle® T7 express *gmd::kan* produced better results respect to $GlcNAcMan_3GlcNAc_2$ synthesis, possibly because of disulfide bond formation which may play a role on MstX:ΔTMGnTI structure and stability.

Given the quantitative limitations inherent to MS, it is still unclear how much of this increase in efficiency is effective or is just related to accumulation of the precursor $ManGlcNAc_2$ (peak at 763 in figure 4.5b) and subsequent reduction in $Man_3GlcNAc_2$ production. Future experiments are planned for glycan quantification and improved comparison of $Man_3GlcNAc_2$ related to expression of MstX:ΔTMGnTI. Accumulation of $ManGlcNAc_2$ is also interesting given the possibility for MstX:ΔTMGnTI to be interfering directly with the enzymatic ac-

tivity of Alg2. It could be speculated that even though MstX:ΔTMGnTI cannot attach a GlcNAc residue to $ManGlcNAc_2$, its affinity towards mannose could make it compete with Alg2 for this substrate. Additional evidence from our colaborators at Glycobia inc. might support this asseveration. They evaluated *in vivo* activity of a *N*-acetylglucosaminyltransferase from *Nicotiana tabacum* (nGnTI), finding that it was active and that $ManGlcNAc_2$ accumulated also in this case.

In conclusion, human GnTI was successfully expressed in *E. coli* after modifications like removal of the *N*-terminal transmembrane domain and conjugation to folding and membrane integration partners. Furthermore, *in vivo* activity was demonstrated, although efficiency is relatively low. Finally, host engineering was applied to increase efficiency by use of oxidizing cytoplasm *E. coli* strains, facilitating formation of disulfide bonds and possibly stabilizing GnTI structure.

CHAPTER 5

OPTIMIZING THE SYNTHESIS OF THE TRIMANNOSE CORE IN *ESCHERICHIA COLI*

5.1 Introduction

After successful transfer of the *N*-liked protein glycosylation machinery from *Campylobacter jejuni* to *E. coli* [173], the idea of engineering bacteria for production of genuine human glycoproteins is becoming closer to reality. Recently, our lab in colaboration with Glycobia inc. developed an *E. coli* based system for the synthesis the eukaryotic *N*-linked glycan common trimannose core ($Man_3GlcNAc_2$) and further transfer of this oligosaccharide to a target glycoprotein [167]. The system relies mainly on recombinant expression of several glycosyltransferases from *Saccharomyces cerevisiae* and an oligosaccharyltransferase (PglB) from *C. jejuni*.

The engineering of eukaryotic-like protein glycosylation in *E. coli* served as proof of concept for the future development of human-like protein glycosylation in bacteria. However, alongside with the synthesis of humanized glycans in *E. coli*, it is also necessary to assess a number of drawbacks in our current system.

Given glycosyltransferases are membrane proteins, their recombinant expression is often difficult, being poor expression and toxicity in the bacterial host the more common problems [16, 43, 50, 100, 113, 175]. Our system is not the exception, and cell growth rate is severely diminished as a consequence of the expression of some of the *S. cerevisiae* glycosyltransferases. On the other hand PglB, which is also a membrane protein, can affect cell growth as well.

One possible way to overcome the cell growth problem focuses on engineering of the host. Previously, the pressure exerted by membrane protein overex-

pression was exploited to isolate spontaneous mutants able to stand to this toxic effect [111]. A similar approach was successfully applied by our collaborators at Glycobia inc. for isolation of mutants which growth rate is less affected by expression of glycosyltransferases.

Interestingly, later studies found that part of the reason why this mutant hosts were more resistant to membrane protein overexpression was because they were able to repress expression at the genetic level [176]. Thereafter, this concept was developed into the restrained expression approach [118], which rationally represses transcription while enhancing overall expression and membrane integration.

A third approach deals with the amino acid sequence of the protein. Directed evolution of glycosyltransferases has been applied to modify their enzymatic activity [166, 183, 182] as well as to enhance natural activity [181]. In our case, we took advantage of this approach for enhancing enzymatic activity of selected glycosyltransferases under specific conditions like expression level and bacterial host. To do so, we took advantage of some interesting features of the $Man_3GlcNAc_2$ synthetic machinery, like display of this glycan on bacterial surface.

In general, gram negative bacteria synthesize specific glycans known as *O*-antigens that are attached to lipid A by WaaL, a relatively promiscuous ligase [168], resulting in the structure known as lipopolysaccharide. Later, lipopolysaccharide is transported to the outer membrane by a 7 protein complex machinery [31, 155] and *O*-antigens are then displayed on cell surface where they are available for interaction with external agents. Given the $Man_3GlcNAc_2$ synthesis is based partially on the *O*-antigen synthetic machinery, it is not surprising that WaaL recognizes this structure and therefore $Man_3GlcNAc_2$ is incor-

porated into the lipopolysaccharide and results displayed on the outer surface of cells. Figure 5.1 presents a schematic of this process.

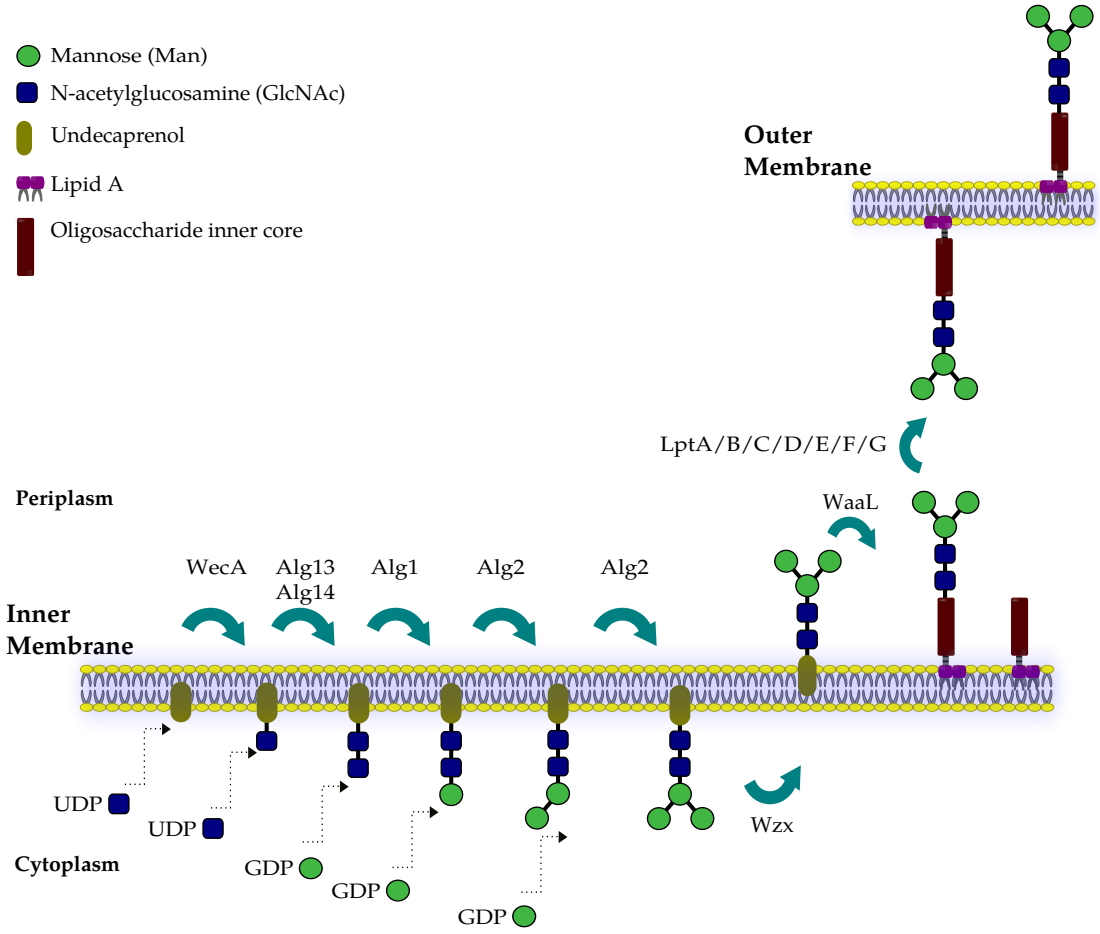

Figure 5.1: *Synthesis and surface display of $Man_3GlcNAc_2$ in E. coli.* In this picture a schematic for glycan synthesis in the inner membrane of *E. coli* is shown. Additionally, interaction of this glycan with native lipopolysaccharide synthesis pathway leads to display of the trimannose core on the outer membrane.

This behavior can be exploited to track $Man_3GlcNAc_2$ synthesis by labeling this glycan using concanavalin A (ConA), a lectin extracted from *Canavalia ensiformis*. ConA binds preferentially to mannose residues in the trimannose core, rendering cells fluorescent when conjugated to a fluorophore like AlexaFluor. Based on this feature, we hypothesized that improvements in the $Man_3GlcNAc_2$ synthetic machinery should be reflected on cells ConA-AlexaFluor fluorescence

and then this phenotype could be used as a selection tool for sorting libraries of mutant glycosyltransferases. Here we describe a protocol for generation of suitable glycosyltransferase libraries, obtained by error-prone PCR, and selection of mutants with enhanced properties based on FACS selection.

5.2 Materials and methods

5.2.1 Bacterial strains and media

Antibiotic selection was maintained at: 100 μg/mL ampicillin (Amp), 25 μg/mL chloramphenicol (Cam) and 50 μg/mL kanamycin (Kan). Luria Bertani (LB) media was used for *E. coli* and protein expression was induced by adding L-arabinose and isopropyl -d-thiogalactoside (IPTG) at 0.2% and 100 mM, respectively. Yeast FY834 was maintained on YPD media and synthetic-defined-Uracil media was used to select or maintain yeast plasmids. E. coli MC4100 (F-araD139 Δ(argF-lac)U169 flbB5301 deoC1 ptsF25 relA1 rbsR22 rpsL150 thiA) was used as the recipient strain for genetic manipulations. To delete the *gmd* gene in MC4100 cells, P1*vir* phage transduction was performed using standard methods and the Keio collection [10].

5.2.2 Plasmid construction

Plasmids were constructed using standard homologous recombination in *S. cerevisiae* as previously described [149]. Briefly, pMWO7-YCG was constructed by first PCR amplifying the genes *alg13*, *alg14*, *alg1*, and *alg2* from *S. cerevisiae* genomic DNA with primers containing appropriate regions of overlap. PCR products and the linearized vector pMWO7 (p15a origin and Cam resistance)

were used to transform yeast strain FY834. Constructs assembled in yeast were electroporated into E. coli for verification via PCR, restriction enzyme digestion and/or sequencing. Plasmids pMWO7-YCG-*Δalg14* and pMWO7-YCG-*Δalg2* were derived from pMWO7-YCG by restriction enzyme digestion for removal of either *alg14* or *alg2* and further re-ligation. Plasmids pTRC99-*alg14* and pTRC99-*alg2* were constructed based on pTRC99 using standard techniques. Diagrams of the plasmids used in this study are shown in figure 5.2. Additionally, for easier manipulation, plasmids based on pMWO7 were digested and re-ligated for removal of the genetic information required for working with yeast. This resulted in a "short" version, ~4000bp smaller than the original.

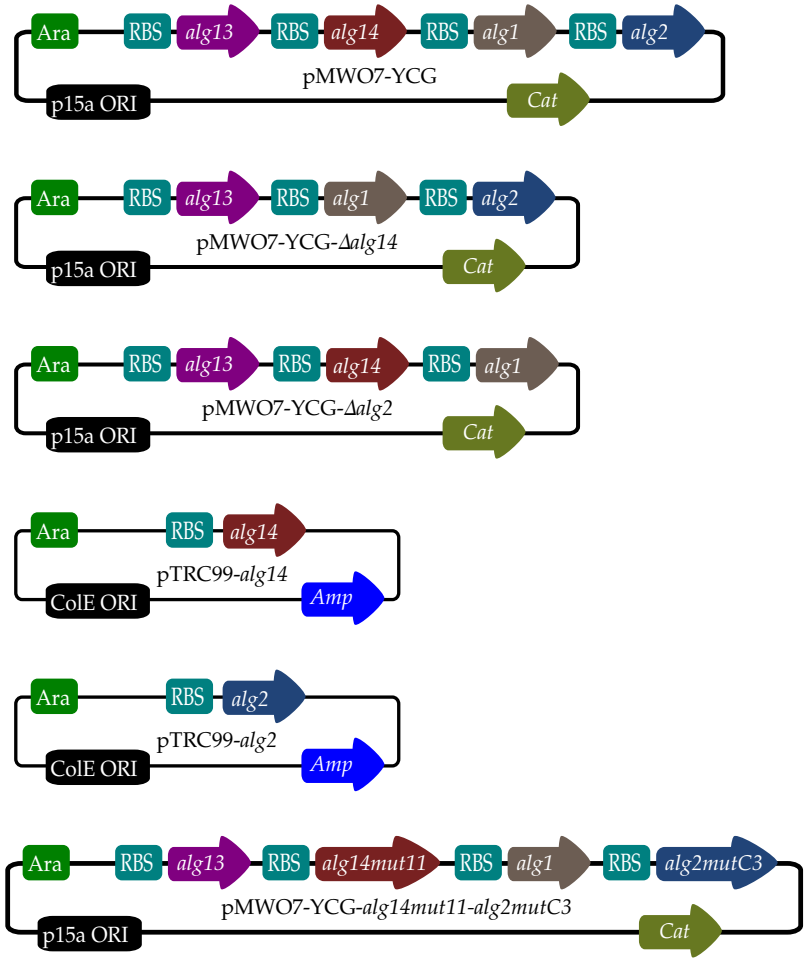

Figure 5.2: *Plasmids used in this study.*

5.2.3 Construction of *alg14* and *alg2* libraries

Genes *alg14* and *alg2* were amplified by error prone PCR using a GeneMorph® II kit (Agilent Technologies) using primers featuring suitable restriction sites. Amplified gene libraries were digested and ligated into pTRC99 using standard techniques. Ligation products were transformed into MC4100 *gmd::kan* carrying either pMWO7-YCG-*Δalg14* or pMWO7-YCG-*Δalg2* and inoculated into 50mL LB supplemented with Cam, Kan, Amp and using L-arabinose and IPTG as inducers. Cells were grown at 30°C under low oxygen conditions for 22h. Flow cytometry was used for sorting cells.

5.2.4 Flow cytometry

E. coli cells were incubated at 30°C overnight in LB supplemented with 0.2% L-arabinose in filled, sealed culture tubes. 1.5mL of the cell suspension was pelleted, washed, and resuspended in (PBS). 2.5 μg/mL *Canavalia ensiformis* Concanavalin A (ConA)-AlexaFluor was added to the samples before incubation in the dark for 30 min at 37°C with constant agitation. 100 μL of each sample were analyzed using a FACSCalibur machine (Becton Dickinson). Propidium iodide was used for labeling dead cells and hence excluding them from the sorting protocol. Selection of highly fluorescent cells was adjusted at ~20 sorted events per second from ~1200 total events per second. After sorting, cells were recovered on a 0.45μm membrane and grown overnight on LB-agar with antibiotics and glucose. Single colonies were used for further analyses.

5.2.5 Fluorophore assisted carbohydrate electrophoresis (FACE)

Glycans were derivatized with 5μL 7amino1,3naphthalenedisulfonic acid (ANDS) 0.15M in 15% acetic acid and 5μL of 1M sodium cyanoborohydride in DMSO, and electrophoresed through a 14% acrylamide gel using 0.01% thorin in 20% glycerol as loading dye, and 1.92M glycine in 0.25M Tris base (pH 8.3) as running buffer at 10mA constant current [56]. Derivatized glycans were detected as fluorescent bands using a UV transilluminator.

5.3 Results

5.3.1 Cells can be sorted based on fluorescent ConA labeling

One important obstacle inherent to our system is the plasmid size. pMWO7-YCG is 11875 base pairs long, and because of this relatively large size, ligation of new inserts in to this plasmid is difficult. Also, transformation efficiency into *E. coli* competent cells decreases linearly with plasmid size [63], another reason why use of pMWO7-YCG for library construction may not be very efficient. One of several possible workarounds to this issue is the use of smaller plasmids for cloning of error-prone PCR libraries. To do so, we modified pMWO7-YCG by eliminating either *alg14* or *alg2*, generating pMWO7-YCG-Δ*alg14* and pMWO7-YCG-Δ*alg2*. These new plasmids permit the use of a second smaller plasmid for complementation of the deleted gene (*alg14* or *alg2*) facilitating the construction of single gene libraries.

This set-up was tested using *alg14* as proof of concept. First, plasmid pMWO7-YCG-Δ*alg14* was co-transformed with plasmid pTRC99-*scFvR4* in MC4100 *gmd::kan* cells as negative control (figure 5.3a). The *scFvR4* gene was

selected because it is slightly larger than *alg14*, facilitating PCR analysis of mixed populations as well as minimizing the potential bias in this analysis. Cell growth did not change significantly when replacing *alg14* for *scFvR4*. We also used pTRC99-*alg14* for complementation of the *alg14* deletion (figure 5.3b), which was successfully tested by ConA-AlexaFluor labeling, obtaining the expected fluorescent phenotype (figure 5.3b).

After showing that *alg14* deletion can be complemented from a second plasmid, a quick PCR based method for cell population analysis was tested. A single set of primers specific for the pTRC99 backbone permitted PCR amplification of either *scFvR4* or *alg14* from pTRC99-*alg14* and pTRC99-*scFvR4*. When these plasmids were used for PCR analysis of cell mixtures carrying either pTRC99-*alg14* or pTRC99-*scFvR4*, a banding pattern consistent with mixture composition is obtained (figure 5.3c). Using this simple method we were able to determine what is the prevalent plasmid in any given mixture. Subsequently, we mixed cells carrying pTRC99-*alg14* or pTRC99-*scFvR4* in a 5% to 95% ratio. After labeling with ConA-AlexaFluor and FACS sorting, 5 random colonies were analyzed finding that 3 of them were carrying pMWO7-YCG-Δ*alg14*, confirming the technical feasibility of sorting cell mixtures in favor of the fluorescent phenotype (figure 5.3d). This experiment served as proof of concept before proceeding to sort an actual *alg14* library.

5.3.2 Alg14 library sorting

MC4100 *gmd::kan* cells carrying pMWO7-YCG-Δ*alg14* and an Alg14 library cloned into pTRC99 were grown overnight under low oxygen conditions for glycan display. After FACS sorting and cell recovery by membrane filtering, 20 colonies were randomly selected for ConA-AlexaFluor labeling and fluores-

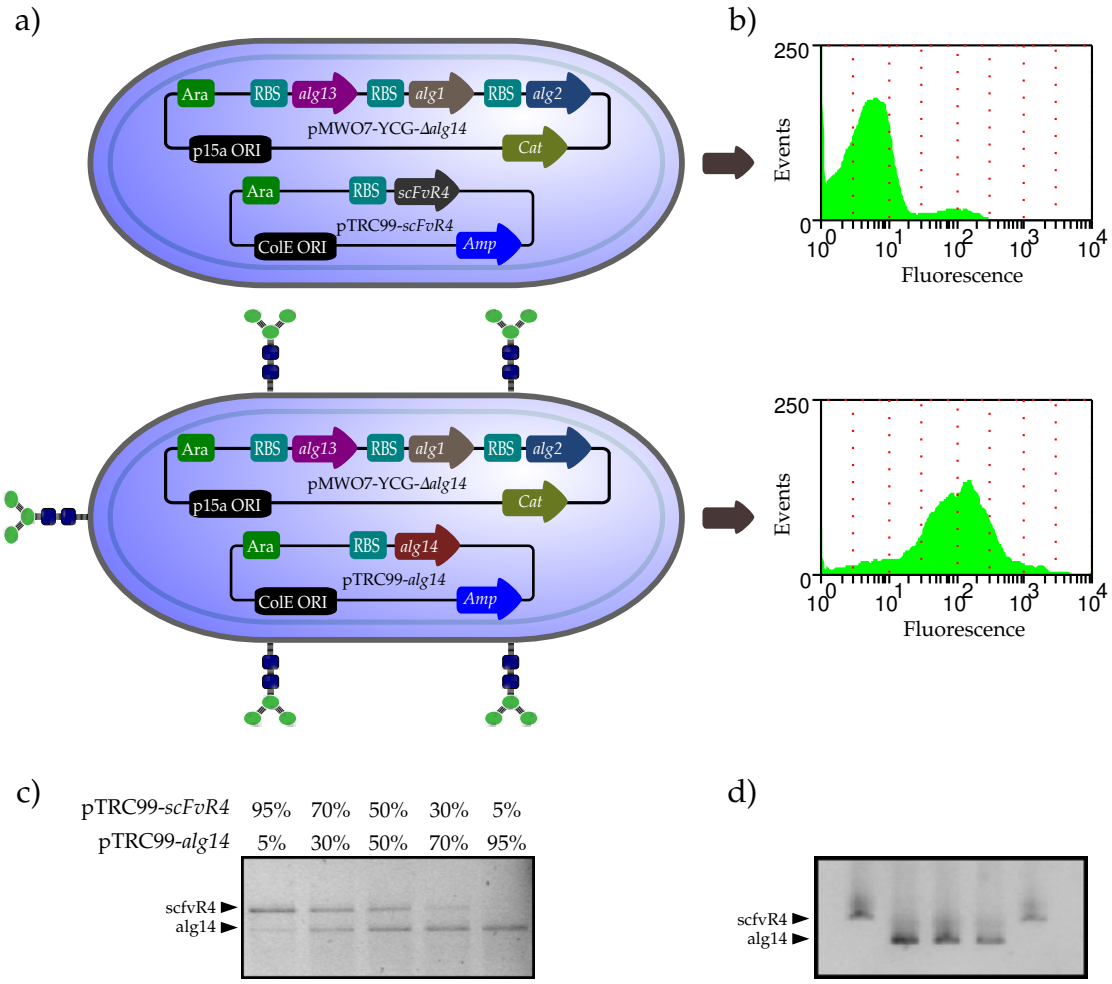

Figure 5.3: *ConA-AlexaFluor fluorescence-based cell sorting: proof of concept*. (a) Schematics of the two plasmids system used for testing the feasibility of sorting cells based ConA-AlexaFluor fluorescence. (b) FACS fluorescence analysis of MC4100 *gmd::kan* cells carrying pMWO7-YCG-*Δalg14* and pTRC99-*scFvR4* or pTRC99-*alg14*. As expected *scFvR4* cannot complement the fluorescent phenotype while Alg4 expression from pTRC99 successfully complemented the fluorescent phenotype. (c) Specific primers for the pTRC99 plasmid were used for PCR-analysis of several mixtures of cells carrying either pTRC99-*scFvR4* or pTRC99-*alg14*. Band intensity correlated with percentage in the mixture. (d) PCR-analysis of a mixture of cells carrying 5% pTRC99-*alg14* and 95% pTRC99-*scFvR4* after FACS sorting and plating on LB-agar with antibiotics. 5 colonies were randomly selected for PCR-analysis resulting in 3 colonies carrying pTRC99-*alg14* (>60%). Based on this result it has been stated that the system can be efficiently sorted based on fluorescence.

cence analysis. Cells showing the highest fluorescence were then compared to cells carrying pMWO7-YCG-*Δalg14* and pTRC99-*alg14* to confirm that selected mutants where showing a better performance. Figure 5.4a presents a comparison of fluorescence among cells carrying YCG-*Δalg14* and the best mutant *alg14* gene (pTRC99-*alg14mut11*), control cells carrying both pMWO7-YCG-*Δalg14* and pTRC99-*alg14*, and a second control population carrying just pMWO7-YCG. Interestingly, cells carrying pTRC99-*alg14mut11*, although more fluorescent than cells carrying pTRC99-*alg14*, are not as bright as cells carrying the single plasmid (pMWO7-YCG).

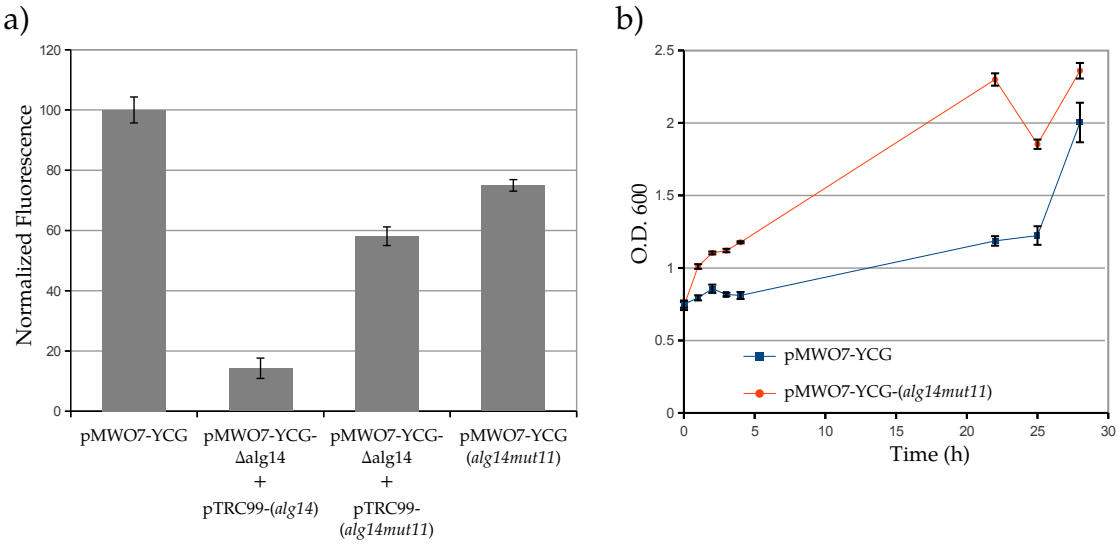

Figure 5.4: *Analysis of Alg14 mutant 11, isolated from sorted library.*. (a) MC4100 *gmd::kan* cells carrying pMWO7-YCG or pMWO7-YCG-*Δalg14* with pTRC99-*alg14* were labeled using ConA-AlexaFluor and compared to cells expressing a mutated version of Alg14 (Alg14mut11). (b) The impact of Alg14 mutations on cell growth was analyzed for cells carrying pMWO7-YCG and pMWO7-YCG-*alg14mut11*

Because of that difference in fluorescence between controls, *alg14mut11* was cloned back into pMWO7-YCG-*Δalg14* to obtain pMWO7-YCG-*alg14mut11*. Surprisingly, cells carrying pMWO7-YCG-*alg14mut11* were not more fluorescent than cells carrying pMWO7-YCG (figure 5.4a). However, when compar-

ing for cell growth, cells carrying pMWO7-YCG-*alg14mut11* grew considerably faster than cells carrying pMWO7-YCG-*alg14* (figure 5.4b), suggesting that *alg14mut11* advantage may be related to reduced toxicity.

5.3.3 Alg2 library sorting

The protocol applied for Alg14 directed evolution was also used for Alg2. In this case pMWO7-YCG-*Δalg2* was complemented using pTRC99-*alg2* and the library was constructed based on this last plasmid. Previously it was shown how overexpression of Alg14 from pTRC99 affects negatively ConA-AlexaFluor fluorescence. On the other hand, we have also observed that overexpression of either Alg13 alongside with Alg14 or Alg2 affects negatively cell growth (data not shown). However, overexpression of Alg2 from pTRC99 does not seem to have any major impact on ConA-AlexaFluor fluorescence. Furthermore, as show in figure 5.5, overexpression of a mutant Alg2 (Alg2mutC3) has a positive effect on ConA-AlexaFluor fluorescence, being it even higher than the single plasmid control pMWO7-YCG, a somewhat unexpected result given the problems observed with Alg14 overexpression.

For a fair comparison with pMWO7-YCG, *alg2mutC3* was cloned back into pMWO7-YCG-*Δalg2* to obtain pMWO7-YCG-*alg2mutC3*. In this case, ConA-AlexaFluor fluorescence for pMWO7-YCG-*alg2mutC3* resulted slightly inferior than for pMWO7-YCG (figure 5.5a). This suggest that Alg2mutC3 increased efficiency may be dependent on overexpression. Nevertheless, overexpression of Alg2mutC3 appears considerably more deleterious than overexpression of wild type Alg2 (figure 5.5b). It is also interesting, as shown in figure 5.6a, that combining the two mutant glycosyltransferases (Alg14mut11 and Alg2mutC3) in a single plasmid system (pMWO7-YCG-*alg14mut11-alg2mutC3*) results in to-

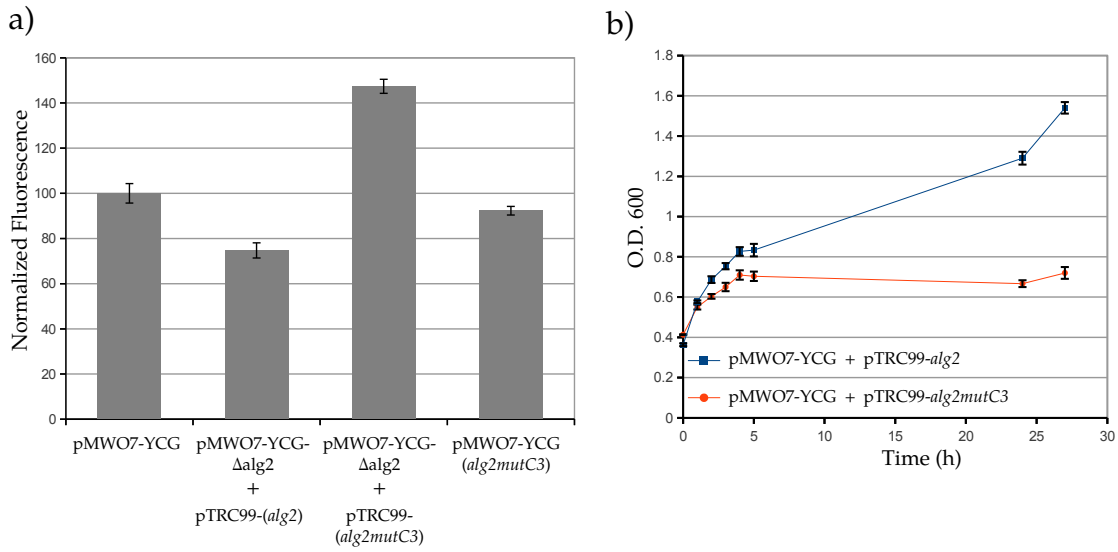

Figure 5.5: *Analysis of Alg2 mutant C3, isolated from sorted library..* (a) MC4100 *gmd::kan* cells carrying pMWO7-YCG or pMWO7-YCG-*Δalg2* with pTRC99-*alg2* were labeled with ConA-AlexaFluor and compared to cells expressing a mutated version of Alg2 (Alg2mutC3). (b) The impact of Alg2 mutations on cell growth was analyzed for cells carrying pMWO7-YCG-*Δalg2* and pTRC99-*alg2* or pTRC99-*alg2mutC3*

tal loss of ConA-AlexaFluor fluorescence. Furthermore, when removing one of the 2 mutations in Alg14mut11[1], there is a minor recovery of ConA-AlexaFluor fluorescence. This result is consistent with a hypothetical interaction between Alg14 and Alg2, but further experimentation is still required.

5.3.4 Analysis of the effect of Alg14 and Alg2 mutations on glycan synthesis

Given ConA-AlexaFluor fluorescence analysis is an indirect method for glycan analysis that only reflects mannose display on cells surface, glycans were extracted from MC4100 *gmd::kan* cells expressing Alg14mut11 or Alg2mutC3.

[1]Removal of these mutations was achieved using the Quick Change protocol using a "short" version of pMWO7-YCG-*alg14mut11-alg2mutC3* where ~4000bp were removed from the original plasmid.

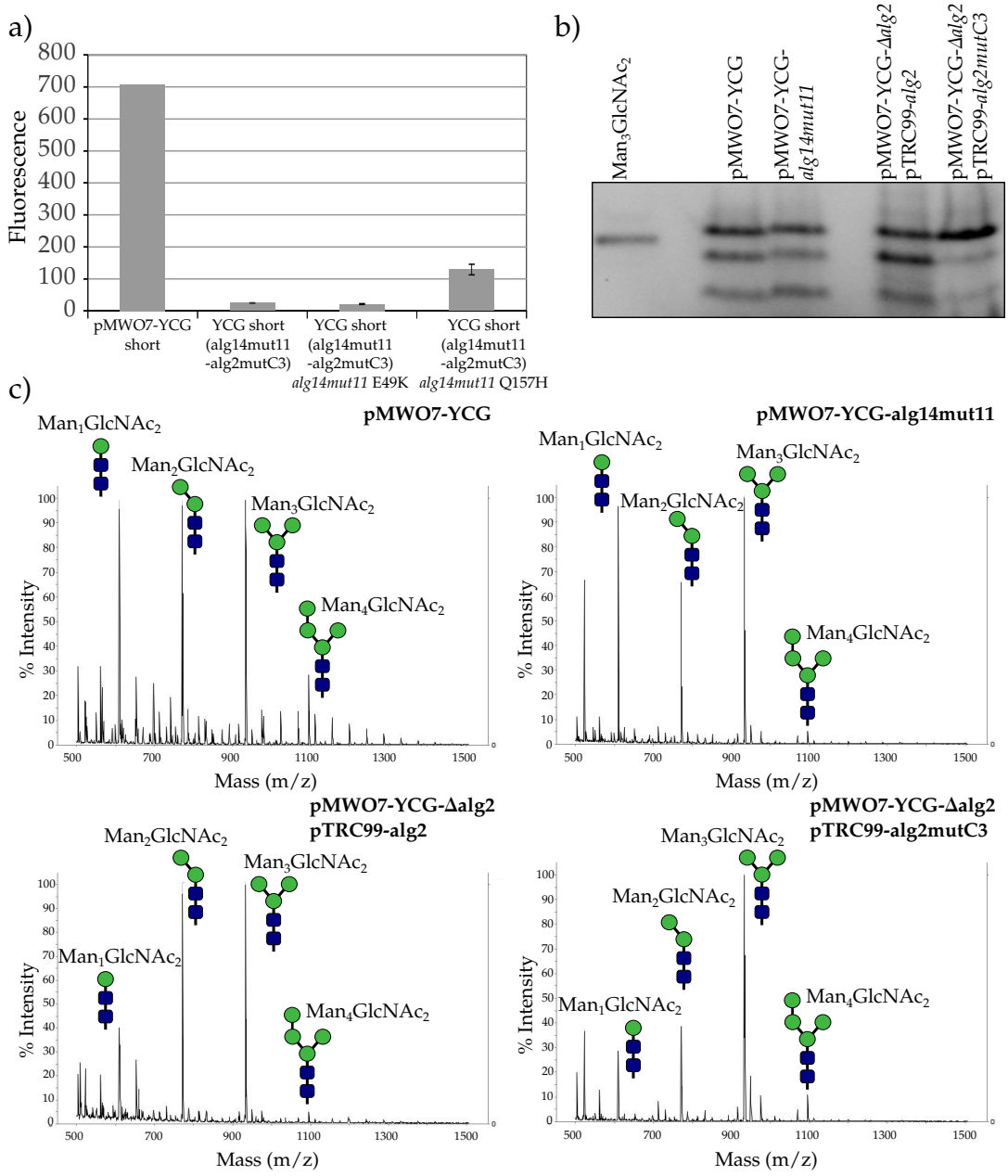

Figure 5.6: *Analysis of the impact of Alg14 and Alg2 mutations on glycan synthesis in MC4100* gmd::kan *cells.*. (a) ConA-AlexaFluor fluorescence analysis of cells carrying both Alg14mut11 and Alg2mutC3 where the 2 mutations in Alg14mut11 have been reverted. Removal of the second mutation results in partial recovery of the fluorescence phenotype. (b) FACE analysis of glycans extracted from MC4100 *gmd::kan* cells carrying either Alg14mut11 or Alg2mutC3. For Alg14mut11 a single plasmid system was used while for Alg2mutC3 a 2 plasmid system was selected. Samples are normalized respect to O.D.600nm. This experiment shows that there is a proportional relation between cell fluorescence and glycan production in the inner membrane. (c) MS analyses corresponding to the samples analyzed by FACE. The analysis suggests that Alg2mutC3 is more efficient than Alg2 wild type respect to $Man_3GlcNAc_2$ synthsis.

For the Alg14mut11 case, a single plasmid system (pMWO7-YCG-*alg14mut11*) was selected given this set-up resulted in higher ConA-AlexaFluor fluorescence. For the Alg2mutC3 case, a double plasmid system (YCG-*Δalg2* with pTRC99-*alg2mutC3*) was used because of similar reasons. Glycans were analyzed using FACE (figure 5.6b) obtaining results that correlate with previous ConA-AlexaFluor fluorescence analyzes, showing that $Man_3GlcNAc_2$ yield is similar for all cases except for cells carrying YCG-*Δalg2* and pTRC99-*alg2mutC3*, where $Man_3GlcNAc_2$ yield seems to be higher. On the other hand, MS analyzes (figure 5.6c) permitted us to compare $Man_3GlcNAc_2$ synthesis respect to $Man_1GlcNAc_2$ and $Man_2GlcNAc_2$ precursors (results are normalized respect to $Man_3GlcNAc_2$) to establish possible differences between wild type and mutated glycosyltransferases.

5.4 Discussion

Sorting results suggest that ConA-AlexaFluor fluorescence-based FACS can select for glycosyltransferases with enhanced properties. Interestingly, Alg14mut11 does not appear to have an enhanced catalytic activity, but instead its negative effect over cell growth is less pronounced than that of the wild type Alg14, suggesting that this mutant may be less toxic. Givent that, it could also be speculated that low ConA-AlexaFluor fluorescence in the double plasmid system (YCG-*Δalg14* with pTRC99-*alg14* or pTRC99-*alg14mut11*) may be also related to Alg14 toxicity and hence Alg14 expression should be restrained to maintain cell viability.

In the Alg2 directed evolution case, the observations are in several ways different. Here, the double plasmid system (YCG-*Δalg2* with pTRC99-*alg2mutC3*)

resulted in a higher ConA-AlexaFluor fluorescence and even when expressing the wild type Alg2 (YCG-*Δalg2* with pTRC99-*alg2*) fluorescence is comparable to pMWO7-YCG. However, when analyzing cell growth, Alg2mutC3 appears to severely affect the system, completely abolishing growth even 24h after induction. Nevertheless, the single plasmid system (pMWO7-YCG-*alg2mutC3*) is only slightly less fluorescent than pMWO7-YCG, which in some way results counter-intuitive, considering that Alg2mutC3 appears to be highly toxic for cells, but its enhanced catalytic activity seems to rely on overexpression.

Respect to our initial hypothesis, an important results is that FACE analysis from extracted glycans correlates satisfactorily with ConA-AlexaFluor fluorescence analysis. When comparing bands corresponding to $Man_3GlcNAc_2$ in figure 5.6b, all samples presented similar band intensity except for cells expressing Alg2mutC3 (YCG-*Δalg2* with pTRC99-*alg2mutC3*), where band intensity was higher. Given the extraction protocol favor isolation of LLOs from the inner membrane, this result serves as evidence for a proportional relation between LLOs synthesis in the inner membrane and glycan display in the outer membrane after transfer to lipid A by WaaL. This would validate ConA-AlexaFluor fluorescence as a suitable method for estimation of LLO synthesis efficiency and glycosyltransferase efficiency, by extension.

On the other hand, MS gives us some clues about the relative activity of mutant glycosyltransferases. First, Alg14mut11 could be indirectly responsible for the accumulation of $Man_1GlcNAc_2$ (figure 5.6c top) given an increase in $GlcNAc_2$ production and assuming Alg1 is more efficient than Alg14 and Alg2[2]. Second, overexpression of Alg2 accounts for depletion of $Man_1GlcNAc_2$, favor-

[2]Overexpression of Alg1 did not show any negative impact on *E. coli* cells growth. Also, C-terminal GFP conjugation analysis results in higher fluorescence for Alg1 in comparison to Alg14 and Alg2 (Data not shown).

ing $Man_2GlcNAc_2$ and $Man_3GlcNAc_2$. In the same way Alg2mutC3 would also deplete $Man_1GlcNAc_2$ but this time favoring $Man_3GlcNAc_2$ mainly. Unfortunately, Alg2mutC3 toxicity limits its usefulness based on the enzymatic advantage.

In general, the complexity of the system makes hard to predict the directed evolution outcome. As shown in this study, using similar set-ups, a less toxic but slightly less efficient glycosyltransferase was obtained and a more toxic but more efficient glycosyltransferase was also obtained. Additionally, expression levels of these glycosyltransferases affect dramatically their behavior, where in the less expected case, activity and toxicity are increased at the same time after directed evolution. Based on these results, a more robust selection system should be based on a single plasmid instead of the current 2 plasmid system. Ideally, the whole operon would be randomly mutated as opposed to single gene mutagenesis, so glycosyltransferases are co-evolved preventing incompatibilities as observed for pMWO7-YCG-*alg14mut11-alg2mutC3* (figure 5.6a).

Given a crystal structure is not available for either Alg14 or Alg2, analysis of mutations is difficult. Alg14mut11 has two mutations: E49K and Q157H. According to Alg14 predicted topology [8] the first mutation occurs just after the *N*-terminal transmembrane domain, changing from a negatively charged residue to a positively charged one, which as suggested in previous studies [14, 98], may affect directionality and membrane insertion and could explain the decreased toxicity. The second mutation, Q157H, happens on a *C*-terminal membrane associated domain. Here, a hydrophobic residue is changed for a positively charged one, which may be disrupting proper association of this domain with the membrane. Notably, this disruption is consistent with partial recovery of ConA-AlexaFluor fluorescence after reverting this mutation in

pMWO7-YCG-*alg14mut11-alg2mutC3* and could be involved in the hypothetical Alg14-Alg2 interaction (figure 5.6a).

On the other hand, 4 mutations were found in Alg2mutC3: A101V, E243K, T285A and a silent mutation at V166 (GTG to GTC). A101V is located just after the first predicted transmembrane domain [81], however analysis of this mutation is difficult given both A and V are hydrophobic residues and, according to codon usage data [119], codons in Alg2 and Alg2mutC3 at this position are both optimal for *E. coli*. E243K and T285 may have an important effect on structure and/or catalytic activity, given these two mutations are located in the main central soluble loop where a negatively charged residue is replaced for a positively charged residue in the first case, and a polar residue is replaced for a hydrophobic residue in the second case. Finally, the silent mutation at V166 is located after the second transmembrane domain, changing from the optimal codon to a sub-optimal one. Theoretically, this could slow down protein synthesis having a positive effect on protein folding and membrane insertion.

In summary, ConA-AlexaFluor labeling can be used for directed evolution of Alg glycosyltransferases in *E. coli*. Mutant glycosyltransferases with enhanced activity or decreased toxicity were obtained by this method. Still, results made evident the importance of co-evolution of this glycosyltransferases to avoid general issues derived from non-compatible mutations. Also, expression levels of this enzymes seems to play an important role on glycan synthesis, a reason why this variable should be included in future evolution experiments like, for example, by mutating ribosome binding sites preceding each gene in the Alg operon.

CHAPTER 6

CONCLUDING REMARKS AND FUTURE WORK

6.1 Conclusions

Synthesis of the eukaryotic trimannose core glycan was achieved for the first time in *E. coli*. This is an important advance not just because this structure has not been observed or engineered in bacteria before, but because the synthetic pathway requires recombinant expression of four eukaryotic glycosyltransferases, three of them membrane proteins. Importantly, the difficulty of recombinant membrane protein expression in *E. coli* has been established, and even when expression is achieved, that does not mean that the protein will be active *in vivo* [16, 43, 50, 100, 113, 175], a reason why synthesis of the trimannose core in this host has been challenging.

Additionally, model proteins were glycosylated *in vivo* by recombinant expression of PglB, a bacterial oligosaccharyltransferase from *C. jejuni*, being this the first time an eukaryotic-like glycosylation machinery has been engineered in bacteria. This glycosylation step requires the native *E. coli* flippases to recognize the lipid-linked trimannose core to make it available for PglB at the periplasmic side of the inner membrane. Additionally, since PglB failed to transfer $GlcNAc_2$ from dolichol to a protein *in vitro* [29], the fact that it can transfer the trimannose core from bactoprenol to a protein *in vivo* (even at a low efficiency) constitutes also an important discovery.

From the engineering point of view, this proof of concept is just the beginning. The engineered glycosylation pathway imposes an important stress on cells, besides not being efficient enough, making optimization an obligatory step. Given this factor, directed evolution was applied on two of the glyco-

syltransferases increasing efficiency in one case and alleviating stress on cells on the other.

Lastly, the trimannose core while good as a starting point, has limited applications, so engineering the synthesis of more complex human-like glycans would open more interesting possibilities. With the synthesis of $Man_5GlcNAc_2$ we were able to transfer the whole cytoplasmic *N*-glycan synthesis pathway from eukaryotes to *E. coli*, at the same time that Alg11 activity was demonstrated for the first time in *E. coli*. On the other hand, we also were able to synthesize $GlcNAcMan_3GlcNAc_2$ by recombinant expression and engineering of human GnTI. This result is perhaps more important since hGnTI constitutes the first step in the synthesis of terminal sialylated glycans, which are of particular interest given their potential therapeutic applications. Furthermore, hGnTI activity was shown in *E. coli* cytoplasm *in vivo* using an LLO as substrate, which is notable taking into account that GnTI is a native Golgi enzyme which uses protein-linked glycans as substrate.

In general, a standardized work flow was established for testing recombinant expression an *in vivo* activity of glycosyltransferases in *E. coli*. This approach permitted the synthesis of the trimannose core and in the future will be crucial for engineering more complex metabolic pathways leading to the synthesis of fully sialylated and highly mannosylated human-like glycans.

6.2 Future work

6.2.1 Extension of the trimannose core

While most of the effort will probably be focused on the synthesis of terminal sialylated glycans, there are additional glycan structures that might also be of interest. For example, fucosylated glycans are recognized for playing a role on a number of processes in mammals like cell adhesion and recognition of blood antigens. They are also involved in some congenital disorders and cancer metastasis [13, 159]. In general, fucose attaches to glycans by α1,2-, α1,3-, α1,4-, and α1,6- linkages [159], however α1,6-fucose is the main glycoform found in human *N*-glycans [112]. Given that, for most practical applications, Fut8 would be required for production of α1,6-fucosylated glycans [75]. Preliminary Fut8[1] expression attempts in *E. coli* have shown relatively good expression levels, although considerable degradation was detected. Importantly, full length Fut8 was also found in the membrane fraction, so since Fut8 is a membrane protein, there are good chances for it to be properly folded and eventually functional *in vivo* (figure 6.1a).

On the other hand, highly mannosylated *N*-glycans are also interesting structures, mainly because $Man_9GlcNAc_2$ is the prevalent glycan displayed on HIV capsule protein gp120 [40]. Having $Man_5GlcNAc_2$ as starting point, addition of the sixth mannose residue is catalyzed by Alg3. This step presents additional challenges when compared to other glycosyltransferases. Alg3 is natively located in the lumen of ER, so instead of GDP-Man it uses dolichol-PP-Man as substrate [37]. To date, no bacterial enzyme has been identified capable of attaching mannose directly to bactoprenol, so synthesis of bactoprenol-PP-

[1] Fut8 from mouse.

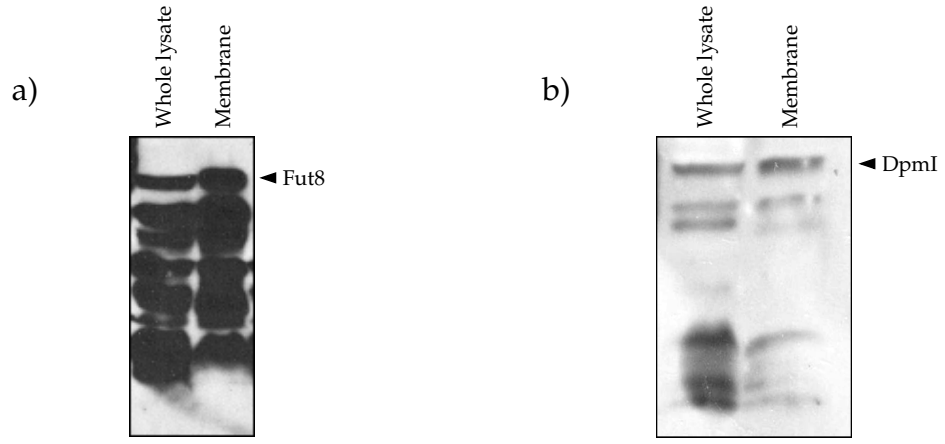

Figure 6.1: *Expression of Fut8 and DpmI in E. coli.* DpmI can be expressed in *E. coli* and it can be found properly localized to the membrane fraction.

Man (as replacement for dolichol-PP-Man in *E. coli*) would require expression of another eukaryotic glycosyltransferase. Dolichol-PP-mannosyltransferases are responsible for the synthesis of dolichol-PP-Man in eukaryotes and usually happen as complexes made of 2 peptide units. However, after cloning and expression in *E. coli*, DpmI was identified as the catalytic subunit in *S. cerevisiae* [105]. Furthermore, *E. coli* expressed DpmI has been shown to transfer mannose to bactoprenol *in vitro* [124], which makes it an excellent candidate for attempting *in vivo* synthesis of bactoprenol-PP-Man. In our hands, DpmI is expressed in *E. coli* and localizes to the membrane fraction, although some degradation is detected (figure 6.1b). Nevertheless, DpmI functionality *in vivo* has not been evaluated yet.

Additionally, Alg3 expression in *E. coli* has been attempted but we still have no immunoblotting data available. Instead, based on previous studies related to membrane protein expression and insertion in the membrane [50, 99, 110], we tested fusion of Alg3 to the *N*-terminus of GFP. As positive control Alg2:GFP was also expressed and fluorescence was evaluated using FACS. While Alg3:GFP was considerably less fluorescent than Alg2:GFP, two peaks

were identified in the histograms, suggesting two *E. coli* populations are present in the sample where one of them could be carrying fluorescent Alg3:GFP (figure 6.2a). Although, while this does not prove that Alg3 is properly expressed, it at least suggests that it could be integrated to the membrane with its *C*-terminus facing the cytoplasm, as expected from computational topology analysis [123]. Finally, *N*-terminal fusions to GlpF and Mstx, as membrane integration partners, were attempted in MC4100 (gmd::kan) resulting in MstX working slightly better. Still, fluorescence is low, possibly because of the relatively large size of the triple fusion MstX:Alg3:GFP (figure 6.2b).

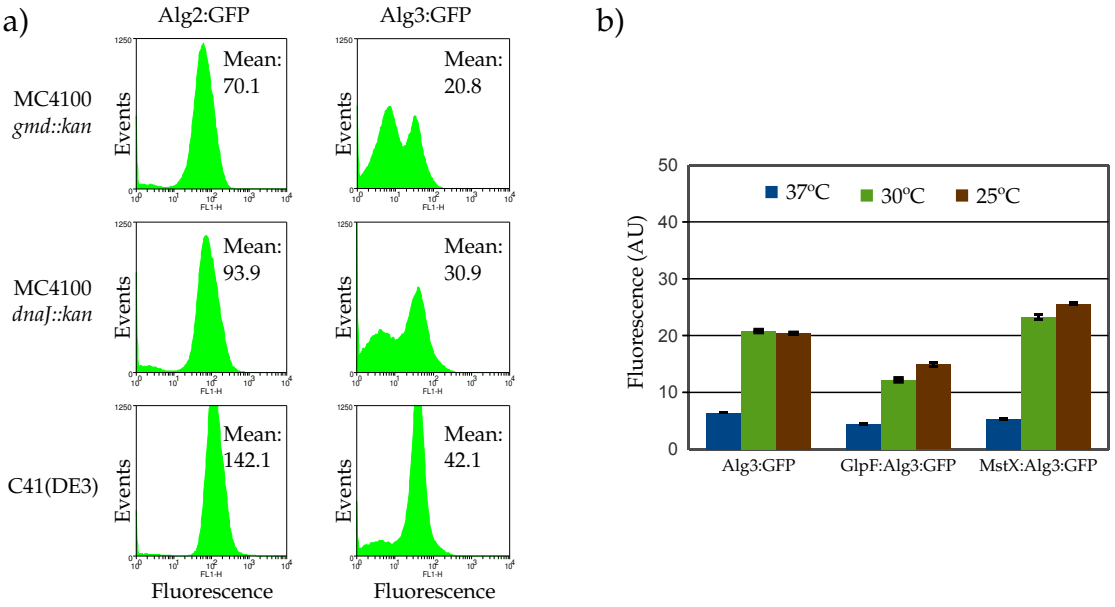

Figure 6.2: *Fluorescence analysis of C-terminal GFP fusions in E. coli.* (a) FACS histograms of Alg2:GFP and Alg3:GFP expressed in several *E. coli* strains. (b) Fluorescence analysis of GlpF:Alg3:GFP and MstX:Alg3:GFP expression at different temperatures. Evidence suggests that Alg3 is poorly expressed in *E.coli* while temperature may have a slight positive effect on expression.

6.2.2 Trimannose core outer membrane display and Outer Membrane Vesicles (OMVs)

An interesting feature in our system is that the trimannose core synthetic pathway can interact with the native *O*-antigen synthetic pathway in *E. coli*. Waal, a ligase that transfer the *O*-antigen from bactoprenol to lipid A, can interact non-specifically with several glycans [168] including the trimannose core. Because of that, the trimannose core reaches the outer membrane after conjugation to lipid A and hence, when synthesized in a hypervesiculating *E. coli* strain, it would be displayed on OMVs outer surface.

It has been suggested that mannosylation can enhance lipososme uptake by macrophages and dendritic cells thanks to interaction with the mannose receptor located in their membrane [32]. As an extension of this observation, we hypothesize that the trimannose core could have a similar effect on bacterial OMVs. On the other hand, a recent study showed that OMVs based vaccination is possible when using antigens located in the luminal space [116], which could be complementary to the classical outer membrane display approach [28]. This in conjunction with the possibility to move plasmids into OMVs [185][41], makes it conceivable the engineering of mannosylated OMVs with potential application as both protein based and DNA based vaccines. A scheme for the system is shown in figure 6.3.

According to Yaron et al. [185], plasmids can get spontaneously into OMVs by a still unknown mechanism. Furthermore, these OMVs can be used for genetic exchange between bacterial cells. In their study, PCR amplification from the OMVs preparation is used as the sole proof of DNA being carried by vesicles. Complementary to this approach, we designed an extended experiment to

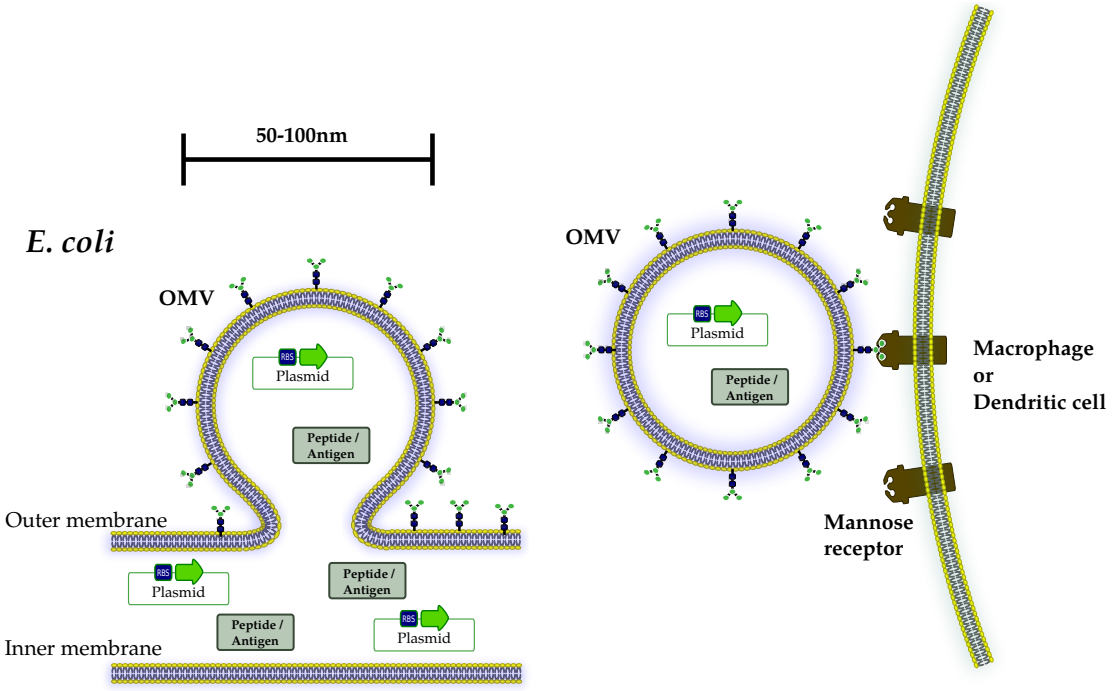

Figure 6.3: *Mannosylated OMVs as vehicle for delivery of protein and DNA based vaccines.*

evaluate if plasmids are effectively inside vesicles instead of being incorporated to the OMVs preparation by any other mean. Briefly, OMVs carrying a plasmid for GFP expression (pTRC-*sstorA:gfp*) were prepared. This preparation was treated with DpnI restriction enzyme[2] and subsequently DNA was amplified by PCR. If pTRC-*sstorA:gfp* is inside vesicles, DpnI should not be able to degrade it and a PCR product will be detected. As a control, miniprepped pTRC-*alg14*[3] was added to the OMVs preparation to confirm DpnI activity. Figure 6.4 shows the results of this experiment.

For all the PCR reactions, standard primers for pTRC plasmids were use so any insert can be amplified using the same set of primers. Lane 1 shows results from PCR amplification directly from OMVs preparation, as described in Yaron et al. [185], a band corresponding to *sstorA:gfp* size was detected. For

[2]DpnI digests methylated bacterial DNA.
[3]This plasmid was randomly selected only as positive control for DpnI activity

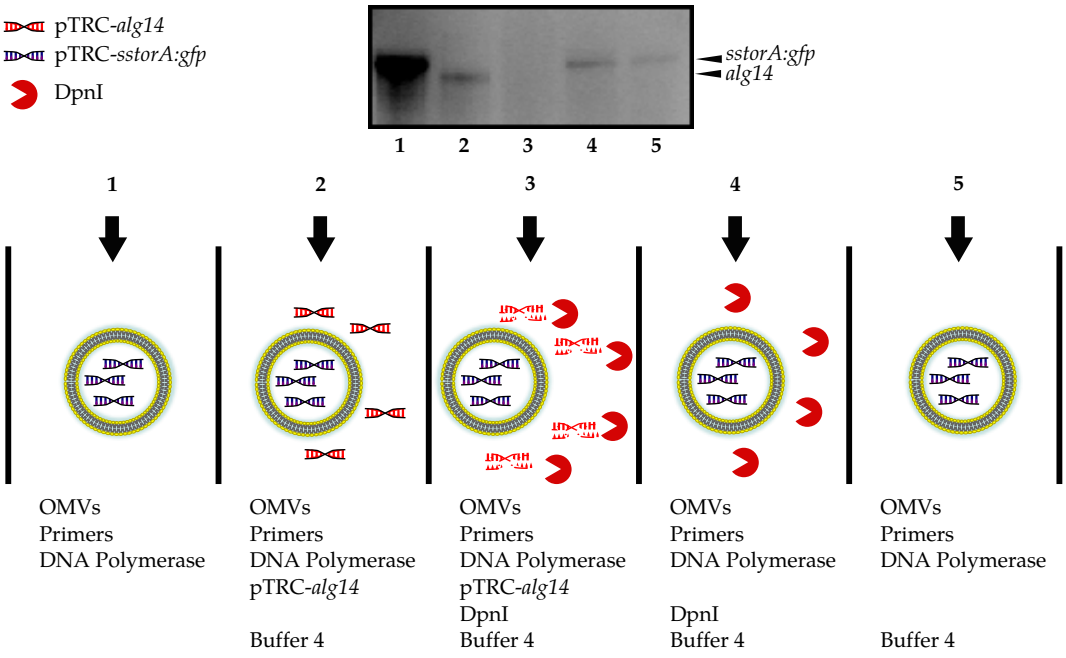

Figure 6.4: *Analysis of plasmid incorporation to the luminal space in OMVs.*

the experiment in lane 2, pTRC-*alg14* was added to the preparation as well as buffer 4. Buffer 4 is necessary for DpnI digestion so it was added as control even in the absence of DpnI. Bands corresponding to *sstorA:gfp* and *alg14* were expected. However, only an *alg14* band was detected. A possible explanation is that since pTRC-*alg14* is outside vesicles, it is more readily accessible to DNA polymerase, so *alg14* amplification overcompetes *sstorA:gfp* amplification and this second product is not detected. After DpnI digestion no bands were detected at all (lane 3). The possible reason for this is that DpnI digestion products from pTRC-*alg14* can still anneal to primers so they interfere with *sstorA:gfp* amplification. For this explanation to be valid, the same experiment without addition of pTRC-*alg14* should lead to *sstorA::gfp* amplification even in the presence of DpnI, as confirmed in lane 4. Finally, lane 5 shows *sstorA:gfp* amplification only in the presence of Buffer 4 and OMVs. Buffer 4 seems to be affecting DNA polymerase performance, which can be inferred from lane 5 result as well as

from comparison of lane 1 to all the other lanes. According to this results, we believe pTRC-*sstorA:gfp* is being incorporated to the inside of OMVs.

Given our proposed system also requires targeting glycans to OMVs surface, as well as proteins to the lumen, we transformed JC8031 *gmd::kan* cells with pMWO7-YCG for synthesis of the trimannose core and pTRC-*sstorA:gfp* for synthesis of ssTorA:GFP and targeting to *E. coli* periplasm (and hence OMVs lumen) thanks to TorA signal sequence (ssTorA). As shown in figure 6.5a, we tested OMVs fluorescence to determine GFP presence in the lumen. First we tested using low oxygen conditions, given MC4100 *gmd::kan* only display the trimannose core on its surface under this conditions, however OMVs did not show any fluorescence. Interestingly, under aerobic conditions, OMVs are fluorescent suggesting that functional GFP is present in the lumen. Additionaly. when probed with ConA-AlexaFluor, OMVs fluorescence increases, probably because the trimannose core is present on the surface as well.

Protein targeting to OMvs was also assessed by western blot analysis. We tested targeting of ssTorA:GFP, ssDsbA:HlyA[4] and ssTorA:HlyA. Figure 6.5b shows that in this particular case ssTorA appears to be very efficient at localizing proteins to OMVs.

Delivery of sstTorA:GFP to mouse macrophages (ATCC J774A.1) *in vitro* was also attempted. Results suggest that OMVs were internalized by macrophages and GFP was hence detected by immunoblotting. Nevertheless, J774A.1 lacks a mannose receptor, a reason why there is no apparent difference between mannosylated and non-mannosylated OMVs (figure 6.5b).

Based on this preliminary data, we consider mannosylated OMVs have potential as a vehicle for delivery of DNA and proteins into macrophages and

[4]HlyA from *Listeria monocytogenes* opens pores in the endosome membrane of macrophages facilitating cell escape or material delivery to the cytoplasm [92, 109].

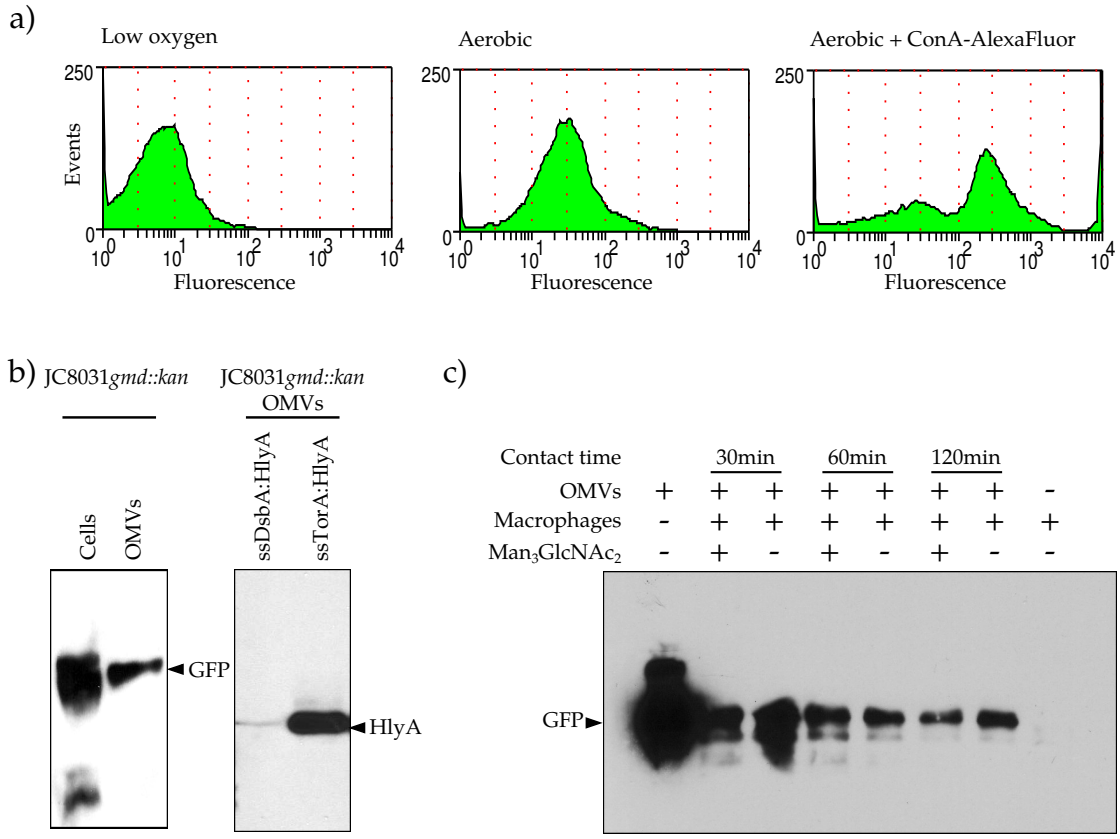

Figure 6.5: *OMVs glycan display and protein targeting to the lumen.* (a) FACS analysis of OMVs from JC8031 *gmd::kan* cells expressing ssTorA:GFP and the trimannose core. (b) Western blot analysis of proteins targeted to the periplasmic space for incorporation into OMVs. (c) OMV based delivery of ssTorA:GFP to mouse macrphages *in vitro*.

dendritic cells, nevertheless further experimentation is still required.

APPENDIX A

OPTIMIZED DNA SEQUENCES

A.1 Codon optimized *pglB*$_{Cj}$

```
   1 ATGCTGAAAA AAGAGTATCT GAAAAACCCT TATCTGGTGC
  41 TGTTCGCCAT GATTGTTCTG GCCTATGTGT TTAGCGTGTT
  81 CTGCCGCTTC TATTGGGTAT GGTGGGCATC CGAATTCAAC
 121 GAGTATTTCT TCAACAACCA GCTGATGATC ATTAGCAATG
 161 ACGGCTATGC CTTTGCCGAA GGTGCTCGTG ATATGATTGC
 201 CGGCTTTCAC CAACCGAACG ATCTGTCCTA TTACGGTAGC
 241 AGCCTGTCTA CACTGACTTA TTGGCTGTAT AAAATCACCC
 281 CGTTCTCATT TGAGAGCATT ATCCTGTATA TGTCCACGTT
 321 TCTGAGCAGT CTGGTAGTAA TCCCGATTAT TCTGCTGGCC
 361 AACGAGTATA AACGCCCGCT GATGGGCTTT GTCGCCGCTC
 401 TGCTGGCAAG TGTGGCAAAT TCGTATTATA ACCGTACCAT
 441 GTCGGGCTAT TACGATACCG ATATGCTGGT CATCGTACTG
 481 CCGATGTTCA TTCTGTTCTT TATGGTCCGT ATGATTCTGA
 521 AAAAAGATTT CTTTAGCCTG ATCGCTCTGC CTCTGTTTAT
 561 TGGCATCTAT CTGTGGTGGT ATCCGTCGTC GTATACCCTG
 601 AATGTTGCCC TGATTGGGCT GTTTCTGATC TATACGCTGA
 641 TCTTCCACCG TAAAGAGAAA ATTTTCTATA TCGCCGTGAT
 681 CCTGTCTAGT CTGACCCTGA GCAATATTGC CTGGTTCTAT
 721 CAGTCAGCCA TCATCGTCAT CCTGTTTGCC CTGTTCGCTC
 761 TGGAACAAAA ACGCCTGAAC TTCATGATTA TTGGCATCCT
 801 GGGTAGCGCT ACGCTGATCT TCCTGATTCT GTCTGGTGGT
 841 GTGGATCCTA TTCTGTATCA ACTGAAATTC TATATTTTCC
```

881 GCTCCGATGA ATCCGCTAAC CTGACACAGG GGTTCATGTA

921 TTTCAACGTC AACCAAACCA TCCAAGAGGT CGAGAATGTC

961 GATTTCTCCG AGTTTATGCG TCGCATTAGT GGCTCTGAGA

1001 TTGTATTCCT GTTCTCACTG TTTGGGTTTG TGTGGCTGCT

1041 GCGTAAACAC AAATCAATGA TTATGGCGCT GCCGATTCTG

1081 GTTCTGGGAT TTCTGGCACT GAAAGGTGGT CTGCGCTTTA

1121 CCATCTATAG CGTTCCGGTA ATGGCACTGG GCTTTGGCTT

1161 TCTGCTGTCC GAGTTCAAAG CAATTCTGGT CAAAAAATAT

1201 TCCCAACTGA CCTCGAATGT GTGTATTGTT TTCGCTACGA

1241 TCCTGACGCT GGCACCTGTT TTTATCCACA TTTATAACTA

1281 TAAAGCACCG ACGGTCTTTT CCCAAAATGA AGCCTCACTG

1321 CTGAATCAAC TGAAAAACAT TGCCAACCGT GAGGACTATG

1361 TGGTAACCTG GTGGGACTAT GGTTATCCGG TTCGCTATTA

1401 TTCCGACGTG AAAACCCTGG TTGATGGTGG TAAACATCTG

1441 GGCAAAGACA ACTTTTTCCC GAGCTTCTCA CTGAGCAAAG

1481 ATGAGCAGGC AGCCGCTAAC ATGGCTCGTC TGAGCGTCGA

1521 GTATACCGAA AAAAGCTTCT ATGCTCCACA AAACGATATC

1561 CTGAAAAGTG ACATCCTGCA GGCCATGATG AAAGACTATA

1601 ACCAGAGCAA CGTCGACCTG TTCCTGGCAT CACTGAGTAA

1641 ACCTGACTTC AAAATCGATA CTCCAAAAAC TCGTGACATT

1681 TATCTGTATA TGCCGGCTCG TATGAGTCTG ATCTTCTCCA

1721 CTGTTGCCTC GTTCTCGTTC ATTAACCTGG ATACGGGTGT

1761 TCTGGACAAA CCGTTCACCT TTTCAACCGC CTATCCGCTG

1801 GACGTGAAAA ATGGCGAGAT CTATCTGAGC AATGGCGTTG

1841 TGCTGTCGGA CGATTTCCGT TCCTTCAAAA TTGGGGACAA

1881 CGTCGTGAGC GTTAACAGTA TTGTCGAGAT TAACAGTATC

```
1921 AAACAAGGCG AGTACAAAAT CACTCCTATC GACGATAAAG
1961 CTCAATTCTA TATCTTCTAT CTGAAAGACT CCGCTATTCC
2001 GTATGCTCAA TTCATTCTGA TGGACAAAAC GATGTTCAAC
2041 TCCGCCTATG TCCAAATGTT CTTCCTGGGC AACTATGACA
2081 AAAACCTGTT CGATCTGGTC ATTAACAGTC GTGACGCCAA
2121 AGTGTTCAAA CTGAAAATCT GA
```

A.2 Codon optimized *pglB*$_{Cl}$

```
   1 ATGAAACTGC AACAGAACTT TACCGATAAC AACTCCATCA
  41 AATATACCTG TATCCTGATC CTGATCGCCT TTGCCTTTAG
  81 TGTGCTGTGT CGCCTGTATT GGGTAGCATG GGCATCCGAA
 121 TTCTATGAGT TTTTCTTCAA CGACCAGCTG ATGATTACCA
 161 CCAACGATGG TTATGCCTTC GCTGAGGGTG CTCGTGATAT
 201 GATTGCCGGC TTCCACCAAC CGAATGATCT GTCCTATTTC
 241 GGCAGCAGTC TGAGTACACT GACATATTGG CTGTATAGCA
 281 TCCTGCCTTT CTCGTTTGAA AGCATTATCC TGTATATGAG
 321 CGCCTTCTTT GCTTCTCTGA TTGTTGTCCC GATTATTCTG
 361 ATCGCTCGTG AGTATAAACT GACCACCTAT GGCTTCATTG
 401 CCGCTCTGCT GGGTTCAATT GCTAACTCGT ATTATAACCG
 441 TACCATGTCG GGCTACTATG ACACTGATAT GCTGGTTCTG
 481 GTTCTGCCAA TGCTGATTCT GCTGACCTTT ATTCGTCTGA
 521 CTATTAACAA AGACATCTTC ACCCTGCTGC TGTCACCGGT
 561 TTTTATCATG ATTTATCTGT GGTGGTATCC GTCCTCTTAT
 601 AGCCTGAATT TCGCCATGAT CGGGCTGTTT GGTCTGTATA
 641 CCCTGGTGTT CCACCGTAAA GAGAAAATCT TCTATCTGAC
```

681 GATCGCCCTG ATGATTATTG CCCTGTCTAT GCTGGCCTGG

721 CAGTATAAAC TGGCCCTGAT TGTTCTGCTG TTTGCCATCT

761 TCGCCTTTAA AGAGGAAAAA ATCAACTTCT ATATGATCTG

801 GGCACTGATC TTCATTAGCA TCCTGATCCT GCATCTGTCT

841 GGAGGTCTGG ATCCAGTACT GTATCAACTG AAATTCTATG

881 TGTTCAAAGC CTCCGATGTT CAAAATCTGA AAGACGCCGC

921 CTTCATGTAT TTTAACGTGA ACGAAACCAT TATGGAGGTG

961 AATACCATTG ATCCGGAAGT CTTTATGCAG CGCATTAGTA

1001 GCAGTGTTCT GGTCTTTATC CTGAGCTTCA TCGGGTTCAT

1041 TCTGCTGTGT AAAGATCACA AAAGCATGCT GCTGGCTCTG

1081 CCTATGCTGG CACTGGGTTT TATGGCTCTG CGTGCGGGTC

1121 TGCGTTTTAC CATTTATGCC GTGCCTGTTA TGGCTCTGGG

1161 TTTTGGCTAT TTCCTGTATG CCTTTTTCAA CTTCCTGGAA

1201 AAAAAACAAA TCAAACTGAG CCTGCGCAAC AAAAACATTC

1241 TGCTGATTCT GATTGCCTTC TTTAGCATTT CTCCGGCACT

1281 GATGCACATC TATTATTATA AAAGCTCCAC CGTGTTTACC

1321 AGCTATGAGG CGTCAATTCT GAATGACCTG AAAAACAAAG

1361 CCCAGCGTGA GGATTATGTT GTGGCATGGT GGGACTATGG

1401 ATATCCGATT CGCTATTATT CCGACGTGAA AACCCTGATC

1441 GATGGCGGAA AACATCTGGG TAAAGACAAC TTCTTCAGCA

1481 GCTTTGTGCT GAGTAAAGAA CAAATTCCGG CAGCGAATAT

1521 GGCTCGTCTG AGCGTGGAGT ATACCGAAAA ATCTTTTAAA

1561 GAAAACTATC CGGACGTGCT GAAAGCCATG GTCAAAGACT

1601 ATAACAAAAC CTCGGCGAAA GACTTTCTGG AGTCCCTGAA

1641 CGACAAAGAT TTCAAATTCG ACACCAACAA AACACGTGAC

1681 GTGTATATCT ATATGCCGTA TCGTATGCTG CGTATCATGC

```
1721 CTGTAGTAGC CCAATTTGCC AACACGAACC CTGACAATGG
1761 GGAGCAAGAG AAATCGCTGT TCTTCTCACA GGCAAACGCC
1801 ATTGCCCAAG ACAAAACGAC CGGTAGCGTT ATGCTGGATA
1841 ACGGTGTGGA GATTATCAAC GACTTTCGTG CCCTGAAAGT
1881 GGAGGGAGCC TCTATTCCAC TGAAAGCGTT CGTCGACATT
1921 GAGAGTATCA CCAATGGAAA ATTCTATTAT AACGAAATCG
1961 ACTCGAAAGC CCAAATCTAT CTGCTGTTCC TGCGCGAGTA
2001 TAAAAGTTTC GTGATCCTGG ATGAAAGCCT GTATAACAGC
2041 TCGTATATTC AAATGTTCCT GCTGAACCAG TATGACCAGG
2081 ACCTGTTTGA GCAAATCACT AACGACACCC GTGCCAAAAT
2121 CTATCGCCTG AAACGCTGA
```

A.3 *alg14mut11*

```
   1 ATGAAAACGG CCTACTTGGC GTCATTGGTG CTCATCGTAT
  41 CGACAGCATA TGTTATTAGG TTGATAGCGA TTCTGCCTTT
  81 TTTCCACACT CAAGCAGGTA CAGAAAAGGA TACGAAAGAT
 121 GGAGTTAACC TACTGAAAAT ACGAGAATCG TCAAAGAAAC
 161 CGCTCAAGAT TTTTGTATTC TTAGGATCGG GAGGTCATAC
 201 TGGTGAAATG ATCCGTCTTC TAGAAAATTA CCAGGATCTT
 241 TTACTGGGTA AGTCGATTGT GTACTTGGGT TATTCTGATG
 281 AGGCTTCCAG GCAAAGATTC GCCCACTTTA TAAAAAAATT
 321 TGGTCATTGC AAAGTAAAAT ACTATGAATT CATGAAAGCT
 361 AGGGAAGTTA AAGCGACTCT CCTACAAAGT GTAAAGACCA
 401 TCATTGGAAC GTTGGTACAA TCTTTTGTGC ACGTGGTTAG
 441 AATCAGATTT GCTATGTGTG GTTCCCCTCA GCTGTTTTTA
```

```
481 TTGAATGGGC CTGGAACATG CTGTATAATA TCCTTTTGGT
521 TGAAAATTAT GGAACTTCTT TTGCCCCTGT TGGGTTCCTC
561 CCATATAGTT TATGTAGAAT CGCTGGCAAG GATTAATACT
601 CCTAGTCTGA CCGGAAAAAT ATTATATTGG GTAGTGGATG
641 AATTCATTGT CCAGTGGCAA GAATTGAGGG ACAATTATTT
681 ACCAAGATCC AAGTGGTTCG GCATCCTTGT TTAA
```

A.4 *alg2mutC3*

```
  1 ATGATTGAAA AGGATAAAAG AACGATTGCT TTTATTCATC
 41 CAGACCTAGG TATTGGGGGC GCTGAAAGGT TAGTCGTCGA
 81 TGCAGCATTA GGTCTACAGC AACAAGGACA TAGTGTAATC
121 ATCTATACTA GTCACTGTGA TAAATCACAT TGTTTCGAAG
161 AAGTTAAAAA CGGCCAATTA AAAGTCGAAG TTTATGGTGA
201 TTTTTTACCG ACAAACTTTT TGGGTCGTTT TTTTATTGTT
241 TTCGCAACAA TTAGACAGCT TTATTTAGTT ATTCAATTGA
281 TCCTACAGAA AAAAGTGAAT GTGTACCAAT TAATTATCAT
321 TGATCAACTG TCTACATGTA TTCCGCTTCT GCATATCTTT
361 AGTTCTGCCA CTTTGATGTT TTATTGTCAT TTCCCCGACC
401 AATTATTGGC TCAAAGAGCT GGGCTATTGA AGAAAATATA
441 CAGACTACCA TTTGACTTAA TAGAACAGTT TTCCGTGAGT
481 GCTGCCGATA CTGTTGTCGT AAATTCAAAT TTCACTAAGA
521 ATACGTTCCA CCAAACGTTC AAGTATTTAT CCAATGATCC
561 AGACGTCATT TATCCATGCG TGGATTTATC AACAATCGAA
601 ATTGAAGATA TTGACAAGAA ATTTTTCAAA ACAGTGTTTA
641 ACGAAGGCGA TAGATTTTAC CTAAGTATAA ATCGTTTTGA
```

681 GAAAAAAAAG GATGTTGCGC TGGCTATAAA GGCTTTTGCG

721 TTATCTAAAG ATCAAATCAA TGACAACGTT AAGTTAGTTA

761 TTTGCGGTGG TTATGACGAG AGGGTTGCAG AAAATGTGGA

801 GTACTTGAAG GAACTACAGT CTCTGGCCGA TGAATACGAA

841 TTATCCCATA CAGCCATATA CTACCAAGAA ATAAAGCGCG

881 TCTCCGATTT AGAGTCATTC AAAACCAATA ATAGTAAAAT

921 TATATTTTTA ACTTCCATTT CATCATCTCT GAAAGAATTA

961 CTGCTCGAAA GAACCGAAAT GTTATTGTAT ACACCAGCAT

1001 ATGAGCACTT TGGTATTGTT CCTTTAGAAG CCATGAAATT

1041 AGGTAAGCCT GTACTAGCAG TAAACAATGG AGGTCCTTTG

1081 GAGACTATCA AATCTTACGT TGCTGGTGAA AATGAAAGTT

1121 CTGCCACTGG GTGGCTAAAA CCTGCCGTCC CTATTCAATG

1161 GGCTACTGCA ATTGATGAAA GCAGAAAGAT CTTGCAGAAC

1201 GGTTCTGTGA ACTTTGAGAG GAATGGCCCG CTAAGAGTCA

1241 AGAAATACTT TTCTAGGGAA GCAATGACTC AGTCATTTGA

1281 AGAAAACGTC GAGAAAGTCA TATGGAAAGA AAAAAAGTAT

1321 TATCCTTGGG AAATATTCGG TATTTCATTC TCTAATTTTA

1361 TTTTGCATAT GGCATTTATA AAAATTCTAC CCAATAATCC

1401 ATGGCCCTTC CTATTTATGG CCACTTTTAT GGTATTATAT

1441 TTTAAGAACT ACTTATGGGG AATTTACTGG GCATTTGTAT

1481 TCGCTCTCTC CTACCCTTAT GAAGAAATAT AA

BIBLIOGRAPHY

[1] B. Absmanner, V. Schmeiser, M. Kampf, and L. Lehle. Biochemical characterization, membrane association and identification of amino acids essential for the function of alg11 from saccharomyces cerevisiae, an alpha1, 2-mannosyltransferase catalysing two sequential glycosylation steps in the formation of the lipid-linked core oligosaccharide. *Biochem. J*, 426:205217, 2010.

[2] Mehtap Abu-Qarn, Jerry Eichler, and Nathan Sharon. Not just for eukarya anymore: protein glycosylation in bacteria and archaea. *Current Opinion in Structural Biology*, 18(5):544–550, October 2008.

[3] Nicholas J. Agard and Carolyn R. Bertozzi. Chemical approaches to perturb, profile, and perceive glycans. *Accounts of Chemical Research*, 42(6):788–797, June 2009.

[4] Bayan Al-Dabbagh, Dominique Mengin-Lecreulx, and Ahmed Bouhss. Purification and characterization of the bacterial UDP-GlcNAc:Undecaprenyl-Phosphate GlcNAc-1-Phosphate transferase WecA. *J. Bacteriol.*, 190(21):7141–7146, November 2008.

[5] C. Alaimo, I. Catrein, L. Morf, C. L. Marolda, N. Callewaert, M. A. Valvano, M. F. Feldman, and M. Aebi. Two distinct but interchangeable mechanisms for flipping of lipid-linked oligosaccharides. *The EMBO Journal*, 25:967–976, 2006.

[6] Dana C Andersen and Lynne Krummen. Recombinant protein expression for therapeutic applications. *Current Opinion in Biotechnology*, 13(2):117–123, April 2002.

[7] Tsutomu Arakawa, Steven J. Prestrelski, William C. Kenney, and John F. Carpenter. Factors affecting short-term and long-term stabilities of proteins. *Advanced Drug Delivery Reviews*, 10(1):1–28, April 1993.

[8] Nicole Averbeck, Sabine Keppler-Ross, and Neta Dean. Membrane topology of the alg14 endoplasmic reticulum UDP-GlcNAc transferase subunit. *J. Biol. Chem.*, 282(40):29081–29088, October 2007.

[9] David Aviezer, Einat Brill-Almon, Yoseph Shaaltiel, Sharon Hashmueli, Daniel Bartfeld, Sarah Mizrachi, Yael Liberman, Arnold Freeman, Ari

Zimran, and Eithan Galun. A Plant-Derived recombinant human glucocerebrosidase EnzymeA preclinical and phase i investigation. *PLoS ONE*, 4(3):e4792, March 2009.

[10] Tomoya Baba, Takeshi Ara, Miki Hasegawa, Yuki Takai, Yoshiko Okumura, Miki Baba, Kirill A Datsenko, Masaru Tomita, Barry L Wanner, and Hirotada Mori. Construction of escherichia coli k-12 in-frame, single-gene knockout mutants: the keio collection. *Mol Syst Biol*, 2, February 2006.

[11] J. Balzarini. Inhibition of HIV entry by carbohydrate-binding proteins. *Antiviral Research*, 71(2-3):237–247, September 2006.

[12] Jan Balzarini, Katrien O Franois, Kristel Van Laethem, Bart Hoorelbeke, Marleen Renders, Joeri Auwerx, Sandra Liekens, Toshikazu Oki, Yasuhiro Igarashi, and Dominique Schols. Pradimicin s, a highly soluble nonpeptidic Small-Size Carbohydrate-Binding antibiotic, is an Anti-HIV drug lead for both microbicidal and systemic use. *Antimicrobial Agents and Chemotherapy*, 54(4):1425–1435, April 2010.

[13] Daniel J Becker and John B Lowe. Fucose: Biosynthesis and biological function in mammals. *Glycobiology*, 13(7):41R–53R, July 2003.

[14] J P Beltzer, K. Fiedler, C. Fuhrer, I. Geffen, C. Handschin, H P Wessels, and M. Spiess. Charged residues are major determinants of the transmembrane orientation of a Signal-Anchor sequence. *Journal of Biological Chemistry*, 266(2):973–978, January 1991.

[15] Markus Berger, Matthias Kaup, and Vronique Blanchard. Protein glycosylation and its impact on biotechnology. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.

[16] Florent Bernaudat, Annie Frelet-Barrand, Nathalie Pochon, Sbastien Dementin, Patrick Hivin, Sylvain Boutigny, Jean-Baptiste Rioux, Daniel Salvi, Daphn Seigneurin-Berny, Pierre Richaud, Jacques Joyard, David Pignol, Monique Sabaty, Thierry Desnos, Eva Pebay-Peyroula, Elisabeth Darrouzet, Thierry Vernet, and Norbert Rolland. Heterologous expression of membrane proteins: Choosing the appropriate host. *PLoS ONE*, 6(12):e29191, December 2011.

[17] Michael D. Best. Click chemistry and bioorthogonal reactions: Unprecedented selectivity in the labeling of biological molecules. *Biochemistry*, 48(28):6571–6584, July 2009.

[18] Tanja Bickel, Ludwig Lehle, Markus Schwarz, Markus Aebi, and Claude A. Jakob. Biosynthesis of lipid-linked oligosaccharides in saccharomyces cerevisiae: Alg13p and alg14p form a complex required for the formation of GlcNAc2-P-Dolichol. *J. Biol. Chem.*, 280(41):34500–34506, October 2005.

[19] Ola Blixt, Nahid Razi, and Minoru Fukuda. Chemoenzymatic synthesis of glycan libraries. In *Glycobiology*, volume Volume 415, pages 137–153. Academic Press, 2006.

[20] Nathan Blow. Glycobiology: A spoonful of sugar. *Nature*, 457(7229):617–620, January 2009.

[21] Piotr Bobrowicz, Robert C. Davidson, Huijuan Li, Thomas I. Potgieter, Juergen H. Nett, Stephen R. Hamilton, Terrance A. Stadheim, Robert G. Miele, Beata Bobrowicz, Teresa Mitchell, Sebastian Rausch, Eduard Renfer, and Stefan Wildt. Engineering of an artificial glycosylation pathway blocked in core oligosaccharide assembly in the yeast pichia pastoris: production of complex humanized glycoproteins with terminal galactose. *Glycobiology*, 14(9):757–766, September 2004.

[22] Mark A. Breidenbach, Jennifer E. G. Gallagher, David S. King, Brian P. Smart, Peng Wu, and Carolyn R. Bertozzi. Targeted metabolic labeling of yeast n-glycans with unnatural sugars. *Proceedings of the National Academy of Sciences*, 107(9):3988–3993, March 2010.

[23] Anders Broberg, Lennart Kenne, and Marianne Pederson. In-situ identification of major metabolites in the red alga gracilariopsis lemaneiformis using high-resolution magic angle spinning nuclear magnetic resonance spectroscopy. *Planta*, 206(2):300–307, August 1998.

[24] Susan Brooks. Strategies for analysis of the glycosylation of proteins: Current status and future perspectives. *Molecular Biotechnology*, 43(1):76–88, 2009.

[25] Patricie Burda and Markus Aebi. The dolichol pathway of n-linked glycosylation. *Biochimica et Biophysica Acta (BBA) - General Subjects*, 1426(2):239–257, January 1999.

[26] Alexandra Castilho, Pia Gattinger, Josephine Grass, Jakub Jez, Martin Pabst, Friedrich Altmann, Markus Gorfer, Richard Strasser, and Herta Steinkellner. N-Glycosylation engineering of plants for the biosynthesis

of glycoproteins with bisected and branched complex n-glycans. *Glycobiology*, 21(6):813–823, June 2011.

[27] Eda Celik, Adam C Fisher, Cassandra Guarino, Thomas J Mansell, and Matthew P DeLisa. A filamentous phage display system for nlinked glycoproteins. *Protein Science*, 19(10):2006–2013, October 2010.

[28] David J. Chen, Nikolaus Osterrieder, Stephan M. Metzger, Elizabeth Buckles, Anne M. Doody, Matthew P. DeLisa, and David Putnam. Delivery of foreign antigens by engineered outer membrane vesicle vaccines. *Proceedings of the National Academy of Sciences*, 107(7):3099–3104, February 2010.

[29] M.M. Chen, K.J. Glover, and B. Imperiali. From peptide to protein: Comparative analysis of the substrate specificity of N-Linked glycosylation in c. jejuni. *Biochemistry*, 46(18):5579–5585, May 2007.

[30] Yasunori Chiba and Yoshifumi Jigami. Production of humanized glycoproteins in bacteria and yeasts. *Current Opinion in Chemical Biology*, 11(6):670–676, December 2007. PMID: 17936668.

[31] S.-S. Chng, N. Ruiz, G. Chimalakonda, T. J. Silhavy, and D. Kahne. Characterization of the two-protein complex in escherichia coli responsible for lipopolysaccharide assembly at the outer membrane. *Proceedings of the National Academy of Sciences*, 107(12):5363–5368, March 2010.

[32] Sumio Chono, Keita Kaneko, Eri Yamamoto, Kohei Togami, and Kazuhiro Morimoto. Effect of surface-mannose modification on aerosolized liposomal delivery to alveolar macrophages. *Drug Development and Industrial Pharmacy*, 36(1):102–107, January 2010.

[33] J F Cipollo, R B Trimble, J H Chi, Q Yan, and N Dean. The yeast ALG11 gene specifies addition of the terminal alpha 1,2-Man to the Man5GlcNAc2-PP-dolichol n-glycosylation intermediate formed on the cytosolic side of the endoplasmic reticulum. *The Journal of Biological Chemistry*, 276(24):21828–21840, June 2001. PMID: 11278778.

[34] Sylvain Cottaz and Eric Samain. Genetic engineering of escherichia coli for the production of NI,NII-diacetylchitobiose (chitinbiose) and its utilization as a primer for the synthesis of complex carbohydrates. *Metabolic Engineering*, 7(4):311–317, July 2005.

[35] JR Couto, TC Huffaker, and PW Robbins. Cloning and expression in escherichia coli of a yeast mannosyltransferase from the asparagine-linked glycosylation pathway. *J. Biol. Chem.*, 259(1):378–382, January 1984.

[36] Kirill A. Datsenko and Barry L. Wanner. One-step inactivation of chromosomal genes in escherichia coli k-12 using PCR products. *Proceedings of the National Academy of Sciences of the United States of America*, 97(12):6640–6645, June 2000.

[37] Robert C. Davidson, Juergen H. Nett, Eduard Renfer, Huijuan Li, Terrance A. Stadheim, Benton J. Miller, Robert G. Miele, Stephen R. Hamilton, Byung-Kwon Choi, Teresa I. Mitchell, and Stefan Wildt. Functional analysis of the ALG3 gene encoding the Dol-P-Man: Man5GlcNAc2-PP-Dol mannosyltransferase enzyme of p. pastoris. *Glycobiology*, 14(5):399–407, May 2004.

[38] Karen De Pourcq, Kristof De Schutter, and Nico Callewaert. Engineering of glycosylation in yeast and other fungi: current state and perspectives. *Applied Microbiology and Biotechnology*, June 2010.

[39] Karen W. Dehnert, Brendan J. Beahm, Thinh T. Huynh, Jeremy M. Baskin, Scott T. Laughlin, Wei Wang, Peng Wu, Sharon L. Amacher, and Carolyn R. Bertozzi. Metabolic labeling of fucosylated glycans in developing zebrafish. *ACS Chemical Biology*, 0(0), 2011.

[40] Katie J. Doores, Camille Bonomelli, David J. Harvey, Snezana Vasiljevic, Raymond A. Dwek, Dennis R. Burton, Max Crispin, and Christopher N. Scanlan. Envelope glycans of immunodeficiency virions are almost entirely oligomannose antigens. *Proceedings of the National Academy of Sciences*, 107(31):13800 –13805, 2010.

[41] D W Dorward, C F Garon, and R C Judd. Export and intercellular transfer of DNA via membrane blebs of neisseria gonorrhoeae. *Journal of Bacteriology*, 171(5):2499–2505, May 1989. PMID: 2496108 PMCID: 209926.

[42] Scot Dowd and Hiroshi Ishizaki. Microarray based comparison of two escherichia coli O157:H7 lineages. *BMC Microbiology*, 6(1):30, 2006.

[43] David Drew, Linda Frderberg, Louise Baars, and Jan-Willem L. de Gier. Assembly and overexpression of membrane proteins in escherichia coli. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1610(1):3–10, February 2003.

[44] C. Dumon, B. Priem, S.L. Martin, A. Heyraud, C. Bosso, and E. Samain. In vivo fucosylation of lacto-N-neotetraose and lacto-N-neohexaose by heterologous expression of helicobacter pylori alpha-1, 3 fucosyltransferase in engineered escherichia coli. *Glycoconjugate journal*, 18(6):465474, 2001.

[45] Yves Durocher and Michael Butler. Expression systems for therapeutic glycoprotein production. *Current Opinion in Biotechnology*, 20(6):700–707, December 2009.

[46] Jens . Duus, Charlotte H. Gotfredsen, and Klaus Bock. Carbohydrate structural determination by NMR spectroscopy: modern methods and limitations. *Chem. Rev.*, 100(12):4589–4614, 2000.

[47] Hay Dvir and Senyon Choe. Bacterial expression of a eukaryotic membrane protein in fusion to various mistic orthologs. *Protein Expression and Purification*, 68(1):28–33, November 2009.

[48] Clemens Drr, Harald Nothaft, Christian Lizak, Rudi Glockshuber, and Markus Aebi. The escherichia coli glycophage display system. *Glycobiology*, 20(11):1366 –1372, November 2010.

[49] Margherita Fais, Rositsa Karamanska, David A. Russell, and Robert A. Field. Lectin and carbohydrate microarrays: New high-throughput methods for glycoprotein, carbohydrate-binding protein and carbohydrate-active enzyme analysis. *Journal of Cereal Science*, 50(3):306–311, November 2009.

[50] J. Fan, J. Heng, S. Dai, N. Shaw, B. Zhou, B. Huang, Z. He, Y. Wang, T. Jiang, X. Li, et al. An efficient strategy for high throughput screening of recombinant integral membrane protein expression and stability. *Protein Expression and Purification*, 2011.

[51] M. F. Feldman, M. Wacker, M. Hernandez, P. G. Hitchen, C. L. Marolda, M. Kowarik, H. R. Morris, A. Dell, M. A. Valvano, and M. Aebi. Engineering n-linked protein glycosylation with diverse o antigen lipopolysaccharide structures in escherichia coli. *Proceedings of the National Academy of Sciences*, 102(8):3016–3021, 2005.

[52] Adam C. Fisher, Charles H. Haitjema, Cassandra Guarino, Eda Celik, Christine E. Endicott, Craig A. Reading, Judith H. Merritt, A. Celeste Ptak, Sheng Zhang, and Matthew P. DeLisa. Production of secretory and extracellular N-Linked glycoproteins in escherichia coli. *Appl. Environ. Microbiol.*, 77(3):871–881, February 2011.

[53] Jordi Folch, M. Lees, and G. H. Sloane Stanley. A SIMPLE METHOD FOR THE ISOLATION AND PURIFICATION OF TOTAL LIPIDES FROM ANIMAL TISSUES. *Journal of Biological Chemistry*, 226(1):497 –509, May 1957.

[54] Kazuhito Fujiyama, Yoshihiro Ido, Ryo Misaki, Daniel G. Moran, Itaru Yanagihara, Takeshi Honda, Shin-Ichiro Nishimura, Toshiomi Yoshida, and Tatsuji Seki. Human n-acetylglucosaminyltransferase i. expression in escherichia coli as a soluble enzyme, and application as an immobilized enzyme for the chemoenzymatic synthesis of n-linked oligosaccharides. *Journal of Bioscience and Bioengineering*, 92(6):569–574, 2001.

[55] Ningguo Gao and Mark A. Lehrman. Alternative sources of reagents and supplies for fluorophore-assisted carbohydrate electrophoresis (FACE). *Glycobiology*, 13(1):1G–3, January 2003.

[56] Ningguo Gao and Mark A Lehrman. Non-radioactive analysis of lipid-linked oligosaccharide compositions by fluorophore-assisted carbohydrate electrophoresis. *Methods in enzymology*, 415:3–20, 2006. PMID: 17116464.

[57] Xiao-Dong Gao, Satoru Moriyama, Nobuaki Miura, Neta Dean, and Shin-Ichiro Nishimura. Interaction between the c termini of alg13 and alg14 mediates formation of the active UDP-N-acetylglucosamine transferase complex. *J. Biol. Chem.*, 283(47):32534–32541, November 2008.

[58] Xiao-Dong Gao, Akiko Nishikawa, and Neta Dean. Physical interactions between the alg1, alg2, and alg11 mannosyltransferases of the endoplasmic reticulum. *Glycobiology*, 14(6):559–570, June 2004.

[59] Xiao-Dong Gao, Hiroyuki Tachikawa, Takashi Sato, Yoshifumi Jigami, and Neta Dean. Alg14 recruits alg13 to the cytoplasmic face of the endoplasmic reticulum to form a novel bipartite UDP-N-acetylglucosamine transferase required for the second step of N-Linked glycosylation. *J. Biol. Chem.*, 280(43):36254–36262, October 2005.

[60] J B Goldberg, K Hatano, G S Meluleni, and G B Pier. Cloning and surface expression of pseudomonas aeruginosa o antigen in escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America*, 89(22):10716–10720, November 1992.

[61] Charlotte H. Gotfredsen, Morten Grtli, Marianne Willert, Morten Meldal, and Jens Duus. Single-bead structure elucidation. requirements for anal-

ysis of combinatorial solid-phase libraries by nanoprobe MAS-NMR spectroscopy. *J. Chem. Soc., Perkin Trans. 1*, (7):1167–1171, January 2000.

[62] Garima Gupta, Avadhesha Surolia, and Srinivasa-Gopalan Sampathkumar. Lectin microarrays for glycomic analysis. *OMICS: A Journal of Integrative Biology*, 14(4):419–436, August 2010.

[63] Douglas Hanahan. Studies on transformation of escherichia coli with plasmids. *Journal of Molecular Biology*, 166(4):557–580, June 1983.

[64] Susan M. Hancock, Mark D. Vaughan, and Stephen G. Withers. Engineering of glycosidases and glycosyltransferases. *Current Opinion in Chemical Biology*, 10(5):509–519, October 2006.

[65] D.J. Harvey. Matrix-assisted laser desorption/ionization mass spectrometry of carbohydrates. *Mass Spectrometry Reviews*, 18(6):349450, 1999.

[66] Marie-Lyn Hecht, Pierre Stallforth, Daniel Varn Silva, Alexander Adibekian, and Peter H Seeberger. Recent advances in carbohydrate-based vaccines. *Current Opinion in Chemical Biology*, 13(3):354–359, June 2009.

[67] A Helenius and M Aebi. Intracellular functions of n-linked glycans. *Science (New York, N.Y.)*, 291(5512):2364–2369, March 2001. PMID: 11269317.

[68] J. Hirabayashi, A. Kuno, and H. Tateno. Lectin-based structural glycomics: A practical approach to complex glycans. *ELECTROPHORESIS*, 32(10):11181128, 2011.

[69] Jason Hollister, Eckart Grabenhorst, Manfred Nimtz, Harald Conradt, and Donald L. Jarvis. Engineering the protein N-Glycosylation pathway in insect cells for production of biantennary, complex N-Glycans. *Biochemistry*, 41(50):15093–15104, 2002.

[70] Patrick Hossler, Bhanu Chandra Mulukutla, and Wei-Shou Hu. Systems analysis of N-Glycan processing in mammalian cells. *PLoS ONE*, 2(8):e713, 2007.

[71] Ku-Lung Hsu, Kanoelani T Pilobello, and Lara K Mahal. Analyzing the dynamic bacterial glycome with a lectin microarray approach. *Nat Chem Biol*, 2(3):153–157, March 2006.

[72] Jian Hu, Huajun Qin, Fei Philip Gao, and Timothy A Cross. A systematic assessment of mature MBP in membrane protein production: Overexpression, membrane targeting and purification. *Protein Expression and Purification*, June 2011. PMID: 21689756.

[73] T C Huffaker and P W Robbins. Temperature-sensitive yeast mutants deficient in asparagine-linked glycosylation. *Journal of Biological Chemistry*, 257(6):3203–3210, March 1982.

[74] D. Huskens, G. Ferir, K. Vermeire, J. C Kehr, J. Balzarini, E. Dittmann, and D. Schols. Microvirin, a novel α (1, 2)-mannose-specific lectin isolated from microcystis aeruginosa, has comparable anti-HIV-1 activity as cyanovirin-N, but a much higher safety profile. *Journal of Biological Chemistry*, 2010.

[75] Hideyuki Ihara, Yoshitaka Ikeda, Sachiko Toma, Xiangchun Wang, Tadashi Suzuki, Jianguo Gu, Eiji Miyoshi, Tomitake Tsukihara, Koichi Honke, Akio Matsumoto, Atsushi Nakagawa, and Naoyuki Taniguchi. Crystal structure of mammalian 1,6-fucosyltransferase, FUT8. *Glycobiology*, 17(5):455 –466, May 2007.

[76] Julian Ihssen, Michael Kowarik, Sandro Dilettoso, Cyril Tanner, Michael Wacker, and Linda Thny-Meyer. Production of glycoprotein vaccines in escherichia coli. *Microbial Cell Factories*, 9(1):61, August 2010.

[77] Karin Ilg, Elif Yavuz, Carola Maffioli, Bernard Priem, and Markus Aebi. Glycomimicry: Display of the GM3 sugar epitope on escherichia coli and salmonella enterica sv typhimurium. *Glycobiology*, 20(10):1289–1297, October 2010.

[78] Pieter P Jacobs, Steven Geysens, Wouter Vervecken, Roland Contreras, and Nico Callewaert. Engineering complex-type n-glycosylation in pichia pastoris using GlycoSwitch technology. *Nat. Protocols*, 4(1):58–70, December 2008.

[79] Jihye JangLee, Simon J. North, Mark SuttonSmith, David Goldberg, Maria Panico, Howard Morris, Stuart Haslam, Anne Dell, and Minoru Fukuda. Glycomic profiling of cells and tissues by mass spectrometry: Fingerprinting and sequencing methodologies. In *Glycobiology*, volume Volume 415, pages 59–86. Academic Press, 2006.

[80] Eric B. Johansen, Francis C. Szoka, Anthony Zaleski, Michael A. Apicella, and Bradford W. Gibson. Utilizing the o-antigen lipopolysaccha-

ride biosynthesis pathway in escherichia coli to interrogate the substrate specificities of exogenous glycosyltransferase genes in a combinatorial approach. *Glycobiology*, page cwq033, March 2010.

[81] M. Kampf, B. Absmanner, M. Schwarz, and L. Lehle. Biochemical characterization and membrane topology of alg2 from saccharomyces cerevisiae as a bifunctional 1,3- and 1,6-Mannosyltransferase involved in lipid-linked oligosaccharide biosynthesis. *Journal of Biological Chemistry*, 284(18):11900–11912, 2009.

[82] Jaroslav Katrlk, Juraj Scaronvitel, Peter Gemeiner, Tibor Kožr, and Jan Tkac. Glycan and lectin microarrays for glycomics and medicinal applications. *Medicinal Research Reviews*, 30(2):394–418, 2010.

[83] H Kayser, R Zeitler, C Kannicht, D Grunow, R Nuck, and W Reutter. Biosynthesis of a nonphysiological sialic acid in different rat organs, using N-propanoyl-D-hexosamines as precursors. *Journal of Biological Chemistry*, 267(24):16934 –16938, 1992.

[84] John Kelly, Harold Jarrell, Lorna Millar, Luc Tessier, Laura M Fiori, Peter C Lau, Brenda Allan, and Christine M Szymanski. Biosynthesis of the n-linked glycan in campylobacter jejuni and addition onto protein through block transfer. *Journal of bacteriology*, 188(7):2427–34, April 2006. PMID: 16547029.

[85] Mirjam M Klepsch, Jan O Persson, and Jan-Willem L de Gier. Consequences of the overexpression of a eukaryotic membrane protein, the human KDEL receptor, in escherichia coli. *Journal of Molecular Biology*, 407(4):532–542, April 2011. PMID: 21316372.

[86] Maria B. Koenigs, Elizabeth A. Richardson, and Danielle H. Dube. Metabolic profiling of helicobacter pylori glycosylation. *Molecular BioSystems*, 5(9):909, 2009.

[87] Michael Kowarik, Shin Numao, Mario F Feldman, Benjamin L Schulz, Nico Callewaert, Eva Kiermaier, Ina Catrein, and Markus Aebi. N-Linked glycosylation of folded proteins by the bacterial oligosaccharyltransferase. *Science*, 314(5802):1148–1150, November 2006.

[88] Michael Kowarik, N Martin Young, Shin Numao, Benjamin L Schulz, Isabelle Hug, Nico Callewaert, Dominic C Mills, David C Watson, Marcela Hernandez, John F Kelly, Michael Wacker, and Markus Aebi. Defini-

tion of the bacterial n-glycosylation site consensus sequence. *EMBO J*, 25(9):1957–1966, May 2006.

[89] Florian Krammer, Sabine Nakowitsch, Paul Messner, Dieter Palmberger, Boris Ferko, and Reingard Grabherr. Swineorigin pandemic H1N1 influenza viruslike particles produced in insect cells induce hemagglutination inhibiting antibodies in BALB/c mice. *Biotechnology Journal*, 5(1):17–23, January 2010.

[90] Christine Krauth, Marta Fedoryshyn, Christian Schleberger, Andriy Luzhetskyy, and Andreas Bechthold. Engineering a function into a glycosyltransferase. *Chemistry & Biology*, 16(1):28–35, January 2009.

[91] Atsushi Kuno, Noboru Uchiyama, Shiori Koseki-Kuno, Youji Ebe, Seigo Takashima, Masao Yamada, and Jun Hirabayashi. Evanescent-field fluorescence-assisted lectin microarray: a new strategy for glycan profiling. *Nat Meth*, 2(11):851–856, November 2005.

[92] Cheng-Yi Kuo, Shubhra Sinha, Jalal A. Jazayeri, and Colin W. Pouton. A stably engineered, suicidal strain of listeria monocytogenes delivers protein and/or DNA to fully differentiated intestinal epithelial monolayers. *Molecular Pharmaceutics*, 6(4):1052–1061, 2009.

[93] Roger A. Laine. Invited commentary: A calculation of all possible oligosaccharide isomers both branched and linear yields 1.05 1012 structures for a reducing hexasaccharide: the isomer barrier to development of single-method saccharide sequencing or synthesis systems. *Glycobiology*, 4(6):759 –767, December 1994.

[94] S. T Laughlin, N. J Agard, J. M Baskin, I. S Carrico, P. V Chang, A. S Ganguli, M. J Hangauer, A. Lo, J. A Prescher, and C. R Bertozzi. Metabolic labeling of glycans with azido sugars for visualization and glycoproteomics. *Methods in enzymology*, page 230250, 2006.

[95] Scott T. Laughlin and Carolyn R. Bertozzi. Imaging the glycome. *Proceedings of the National Academy of Sciences*, 106(1):12–17, January 2009.

[96] B.R Leeflang, E.J Faber, P Erbel, and J.F.G Vliegenthart. Structure elucidation of glycoprotein glycans and of polysaccharides by NMR spectroscopy. *Journal of Biotechnology*, 77(1):115–122, January 2000.

[97] Mark A Lehrman. Teaching dolichol-linked oligosaccharides more tricks

with alternatives to metabolic radiolabeling. *Glycobiology*, 17(8):75R–85, August 2007.

[98] Mirjam Lerch-Bader, Carolina Lundin, Hyun Kim, IngMarie Nilsson, and Gunnar Von Heijne. Contribution of positively charged flanking residues to the insertion of transmembrane helices into the endoplasmic reticulum. *Proceedings of the National Academy of Sciences*, 105(11):4127–4132, March 2008.

[99] A. J. Link, G. Skretas, E. M. Strauch, N. S. Chari, and G. Georgiou. Efficient production of Membrane-Integrated and Detergent-Soluble g Protein-Coupled receptors in escherichia coli. *Protein Science*, 2008.

[100] AJ LINK and G. GEORGIOU. Advances and challenges in membrane protein expression. *AIChE journal*, 53(4):752–756, 2007.

[101] Xin Liu, David J McNally, Harald Nothaft, Christine M Szymanski, Jean-Robert Brisson, and Jianjun Li. Mass spectrometry-based glycomics strategy for exploring n-linked glycosylation in eukaryotes and bacteria. *Analytical Chemistry*, 78(17):6081–6087, September 2006. PMID: 16944887.

[102] Christian Lizak, Sabina Gerber, Shin Numao, Markus Aebi, and Kaspar P. Locher. X-ray structure of a bacterial oligosaccharyltransferase. *Nature*, 474(7351):350–355, June 2011.

[103] Jishun Lu, Tetsuo Takahashi, Atsuko Ohoka, Kei-Ichi Nakajima, Ryo Hashimoto, Nobuaki Miura, Hiroyuki Tachikawa, and Xiao-Dong Gao. Alg14 organizes the formation of a multiglycosyltransferase complex involved in initiation of Lipid-Linked oligosaccharide biosynthesis. *Glycobiology*, 22(4):504–516, April 2012.

[104] Kai Maass, Ren Ranzinger, Hildegard Geyer, ClausWilhelm von der Lieth, and Rudolf Geyer. Glycopeakfinder de novo composition analysis of glycoconjugates. *PROTEOMICS*, 7(24):4435–4444, December 2007.

[105] Yusuke Maeda and Taroh Kinoshita. Dolichol-phosphate mannose synthase: Structure, function and regulation. *Biochimica et Biophysica Acta (BBA) - General Subjects*, 1780(6):861–868, June 2008.

[106] Paola Marani, Samuel Wagner, Louise Baars, Pierre Genevaux, Jan-Willem De Gier, Ingmarie Nilsson, Rita Casadio, and Gunnar Von Heijne. New escherichia coli outer membrane proteins identified through prediction and experimental verification. *Protein Science*, 15(4):884–889, 2006.

[107] Karina Marino, Jonathan Bones, Jayesh J Kattla, and Pauline M Rudd. A systematic approach to protein glycosylation analysis: a path through the maze. *Nat Chem Biol*, 6(10):713–723, October 2010.

[108] Jamey D. Marth and Prabhjit K. Grewal. Mammalian glycosylation in immunity. *Nature Reviews Immunology*, 8(11):874, November 2008.

[109] E Mathew, G E Hardee, C F Bennett, and K-D Lee. Cytosolic delivery of antisense oligonucleotides by listeriolysin o-containing liposomes. *Gene Ther*, 10(13):1105–1115, 2003.

[110] Karen [1] McLuskey, Mads [2] Gabrielsen, Frank [1] Kroner, Isobel [1] Black, Richard [2] Cogdell, and Neil [1] Isaacs. A protocol for high throughput methods for the expression and purification of inner membrane proteins. *Molecular Membrane Biology*, 25:599–608, December 2008.

[111] Bruno Miroux and John E. Walker. Over-production of proteins in-Escherichia coli: Mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *Journal of Molecular Biology*, 260(3):289–298, July 1996.

[112] Eiji Miyoshi, Naofumi Uozumi, Katsuhisa Noda, Norio Hayashi, Masatsugu Hori, and Naoyuki Taniguchi. Expression of 16 fucosyltransferase in rat tissues and human cancer cell lines. *International Journal of Cancer*, 72(6):1117–1121, September 1997.

[113] Daniel Martinez Molina, Tobias Cornvik, Said Eshaghi, Jesper Z. Haeggstrom, Par Nordlund, and Marina Ignatushchenko Sabet. Engineering membrane protein overproduction in escherichia coli. *Protein Sci*, 17(4):673–680, April 2008.

[114] DN Moothoo and JH Naismith. Concanavalin a distorts the beta-GlcNAc-(1–>2)-Man linkage of beta- GlcNAc-(1–>2)-alpha-Man-(1–>3)-[beta-GlcNAc-(1–>2)-alpha-Man- (1– >6)]-Man upon binding. *Glycobiology*, 8(2):173–181, February 1998.

[115] Willy Morelle and Jean-Claude Michalski. Analysis of protein glycosylation by mass spectrometry. *Nat. Protocols*, 2(7):1585–1602, July 2007.

[116] Maneesha Muralinath, Meta J. Kuehn, Kenneth L. Roland, and Roy Curtiss. Immunization with salmonella enterica serovar Typhimurium-Derived outer membrane vesicles delivering the pneumococcal protein

PspA confers protection against challenge with streptococcus pneumoniae. *Infect. Immun.*, 79(2):887–894, February 2011.

[117] J. H. Naismith and R. A. Field. Structural basis of trimannoside recognition by concanavalin a. *Journal of Biological Chemistry*, 271(2):972, 1996.

[118] Anoop Narayanan, Marc Ridilla, and Dinesh A. Yernool. Restrained expression, a method to overproduce toxic membrane proteins by exploiting operator-repressor interactions. *Protein Science*, pages n/a–n/a, 2010.

[119] Frederick C. Neidhardt, editor. *Escherichia coli and Salmonella: Cellular and Molecular Biology*. ASM Press, 2 edition, May 1996.

[120] Irene Neophytou, Richard Harvey, Jayne Lawrence, Phil Marsh, Barry Panaretou, and David Barlow. Eukaryotic integral membrane protein expression utilizing the escherichia coli glycerol-conducting channel protein (GlpF). *Applied Microbiology and Biotechnology*, 77(2):375–381, November 2007.

[121] Christine Noffz, Sabine Keppler-Ross, and Neta Dean. Hetero-oligomeric interactions between early glycosyltransferases of the dolichol cycle. *Glycobiology*, 19(5):472–478, May 2009.

[122] Mary K O'Reilly, Guofeng Zhang, and Barbara Imperiali. In vitro evidence for the dual function of alg2 and alg11: essential mannosyltransferases in n-linked glycoprotein biosynthesis. *Biochemistry*, 45(31):9593–603, August 2006. PMID: 16878994.

[123] Rafael Oriol, Ivan Martinez-Duncker, Isabelle Chantret, Rosella Mollicone, and Patrice Codogno. Common origin and evolution of glycosyltransferases using Dol-P-monosaccharides as donor substrate. *Mol Biol Evol*, 19(9):1451–1463, September 2002.

[124] P. Orlean, C. Albright, and P. W. Robbins. Cloning and sequencing of the yeast gene for dolichol phosphate mannose synthase, an essential protein. *J. Biol. Chem*, 263:17499–17507, 1988.

[125] Michael G. O'Shea, Michael S. Samuel, Christine M. Konik, and Matthew K. Morell. Fluorophore-assisted carbohydrate electrophoresis (FACE) of oligosaccharides: efficiency of labelling and high-resolution separation. *Carbohydrate Research*, 307(1-2):1–12, February 1998.

[126] Vered Padler-Karavani, Hai Yu, Hongzhi Cao, Harshal Chokhawala, Felix Karp, Nissi Varki, Xi Chen, and Ajit Varki. Diversity in specificity, abundance, and composition of anti-Neu5Gc antibodies in normal humans: Potential implications for disease. *Glycobiology*, 18(10):818–830, October 2008.

[127] Jagroop Pandhal and Phillip C Wright. N-Linked glycoengineering for human therapeutic proteins in bacteria. *Biotechnology Letters*, 32(9):1189–1198, September 2010. PMID: 20449632.

[128] Farnoush Parsaie Nasab, Benjamin L. Schulz, Francisco Gamarro, Armando J. Parodi, and Markus Aebi. All in one: Leishmania major STT3 proteins substitute for the whole oligosaccharyltransferase complex in saccharomyces cerevisiae. *Mol. Biol. Cell*, 19(9):3758–3768, September 2008.

[129] Katharina Paschinger, Alba Hykollari, Ebrahim Razzazi-Fazeli, Pamela Greenwell, David Leitsch, Julia Walochnik, and Iain B. H Wilson. The n-glycans of trichomonas vaginalis contain variable core and antennal modifications. *Glycobiology*, October 2011.

[130] Thomas Nordahl Petersen, Sren Brunak, Gunnar von Heijne, and Henrik Nielsen. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods*, 8(10):785–786, September 2011.

[131] Kanoelani T Pilobello and Lara K Mahal. Deciphering the glycocode: the complexity and analytical challenge of glycomics. *Current Opinion in Chemical Biology*, 11(3):300–305, June 2007.

[132] Daniel Poulain and Thierry Jouault. Candida albicans cell wall glycans, host receptors and responses: elements for a decisive crosstalk. *Current Opinion in Microbiology*, 7(4):342–349, August 2004.

[133] Robert Pretel, Kelley Lennon, Alberto Bird, and Maria A. Kukuruzinska. Expression of the first N-Glycosylation gene in the dolichol pathway, ALG7, is regulated at two major control points in the g1 phase of the saccharomyces cerevisiae cell cycle. *Experimental Cell Research*, 219(2):477–486, August 1995.

[134] Justin M. Prien, Bradley D. Prater, and Steven L. Cockrill. A multi-method approach toward de novo glycan characterization: A man-5 case study. *Glycobiology*, 20(5):629–647, May 2010.

[135] Oriol R, Ye Y, Koren E, and Cooper DK. Carbohydrate antigens of pig tissues reacting with human natural antibodies as potential targets for hyperacute vascular rejection in pig-to-man organ xenotransplantation. *Transplantation*, 56(6):1433, 1993.

[136] Vijay Sai Reddy and V.S.R. Rao. Modes of binding of [alpha](1-2) linked manno-oligosaccharides to concanavalin a. *International Journal of Biological Macromolecules*, 14(4):185–192, August 1992.

[137] L. Revers, I B Wilson, M C Webberley, and S L Flitsch. The potential dolichol recognition sequence of beta-1,4-mannosyltransferase is not required for enzymic activity using phytanyl-pyrophosphoryl-alpha-N,N'-diacetylchitobioside as acceptor. *Biochemical Journal*, 299(Pt 1), April 1994. PMC1138015.

[138] Geoffrey D. Robson, Marilyn G. Wiebe, Bryan Cunliffe, and Anthony P. J. Trinci. Choline- and acetylcholine-induced changes in the morphology of fusarium graminearum: evidence for the involvement of the choline transport system and acetylcholinesterase. *Microbiology*, 141(6):1309 – 1314, June 1995.

[139] Tarmo P. Roosild, Jason Greenwald, Mark Vega, Samantha Castronovo, Roland Riek, and Senyon Choe. NMR structure of mistic, a Membrane-Integrating protein for membrane protein expression. *Science*, 307(5713):1317–1321, February 2005.

[140] Marsha R. Rosner, S. Catherine Hubbard, Raymond J. Ivatt, Phillips W. Robbins, and Victor Ginsburg. [35] n-asparagine-linked oligosaccharides: Biosynthesis of the lipid-linked oligosaccharides. In *Complex Carbohydrates Part D*, volume Volume 83, pages 399–408. Academic Press, 1982.

[141] Anne Ruffing and Rachel Chen. Metabolic engineering of microbes for oligosaccharide and polysaccharide synthesis. *Microbial Cell Factories*, 5(1):25, 2006.

[142] John Samuelson, Sulagna Banerjee, Paula Magnelli, Jike Cui, Daniel J. Kelleher, Reid Gilmore, and Phillips W. Robbins. The diversity of dolichol-linked precursors to asn-linked glycans likely results from secondary loss of sets of glycosyltransferases. *Proceedings of the National Academy of Sciences of the United States of America*, 102(5):1548–1553, February 2005.

[143] M Sarkar, S Pagny, U Unligil, D Joziasse, J Mucha, J Glossl, and

H Schachter. Removal of 106 amino acids from the n-terminus of UDP-GlcNAc :alpha-3-D-mannoside beta-1,2-N-acetylglucosaminyltransferase i does not inactivate the enzyme. *Glycoconjugate Journal*, 15(2):193–197, February 1998.

[144] Clark F Schierle, Mehmet Berkmen, Damon Huber, Carol Kumamoto, Dana Boyd, and Jon Beckwith. The DsbA signal sequence directs efficient, cotranslational export of passenger proteins to the escherichia coli periplasm via the signal recognition particle pathway. *Journal of Bacteriology*, 185(19):5706–5713, October 2003.

[145] M. F Schwartz and T. Soliman. *Expression of soluble, active eukaryotic glycosyltransferases in prokaryotic organisms*. Google Patents, March 2006. US Patent App. 11/388,595.

[146] Flavio Schwarz and Markus Aebi. Mechanisms and principles of n-linked protein glycosylation. *Current Opinion in Structural Biology*, 21(5):576–582, October 2011.

[147] Flavio Schwarz, Wei Huang, Cishan Li, Benjamin L Schulz, Christian Lizak, Alessandro Palumbo, Shin Numao, Dario Neri, Markus Aebi, and Lai-Xi Wang. A combined method for producing homogeneous glycoproteins with eukaryotic n-glycosylation. *Nat Chem Biol*, 6(4):264–266, April 2010.

[148] Flavio Schwarz, Christian Lizak, Yao-Yun Fan, Susanna Fleurkens, Michael Kowarik, and Markus Aebi. Relaxed acceptor site specificity of bacterial oligosaccharyltransferase in vivo. *Glycobiology*, 21(1):45 –54, January 2011.

[149] Robert M. Q. Shanks, Nicky C. Caiazza, Shannon M. Hinsa, Christine M. Toutain, and George A. O'Toole. Saccharomyces cerevisiae-Based molecular tool kit for manipulation of genes from Gram-Negative bacteria. *Appl. Environ. Microbiol.*, 72(7):5027–5036, July 2006.

[150] Douglas M. Sheeley and Vernon N. Reinhold. Structural characterization of carbohydrate sequence, linkage, and branching in a quadrupole ion trap mass spectrometer: neutral oligosaccharides and N-Linked glycans. *Anal. Chem.*, 70(14):3053–3059, 1998.

[151] Angus M. Sinclair and Steve Elliott. Glycoengineering: The effect of glycosylation on the properties of therapeutic proteins. *Journal of Pharmaceutical Sciences*, 94(8):1626–1635, 2005.

[152] M Slifkin and R J Doyle. Lectins and their application to clinical microbiology. *Clin. Microbiol. Rev.*, 3(3):197–218, July 1990.

[153] R. J. Sol, J. A. Rodrguez-Martnez, and K. Griebenow. Modulation of protein biophysical properties by chemical glycosylation: biochemical insights and biomedical implications. *Cellular and Molecular Life Sciences*, 64(16):2133–2152, June 2007.

[154] Ricardo J. Sol and Kai Griebenow. Effects of glycosylation on the stability of protein pharmaceuticals. *Journal of Pharmaceutical Sciences*, 98(4):1223–1245, 2009.

[155] Paola Sperandeo, Fion K. Lau, Andrea Carpentieri, Cristina De Castro, Antonio Molinaro, Gianni Deho, Thomas J. Silhavy, and Alessandra Polissi. Functional analysis of the protein machinery required for transport of lipopolysaccharide to the outer membrane of escherichia coli. *J. Bacteriol.*, 190(13):4460–4469, July 2008.

[156] M J Spiro, R G Spiro, and V D Bhoyroo. Lipid-saccharide intermediates in glycoprotein biosynthesis. i. formation of an oligosaccharide-lipid by thyroid slices and evaluation of its role in protein glycosylation. *The Journal of Biological Chemistry*, 251(20):6400–8, October 1976. PMID: 977578.

[157] Robert Sprung, Animesh Nandi, Yue Chen, Sung Chan Kim, Deb Barma, John R. Falck, and Yingming Zhao. Tagging-via-Substrate strategy for probing O-GlcNAc modified proteins. *Journal of Proteome Research*, 4(3):950–957, June 2005.

[158] C. M. Starr, R. Irene Masada, C. Hague, E. Skop, and J. C. Klock. Fluorophore-assisted carbohydrate electrophoresis in the separation, analysis, and sequencing of carbohydrates. *Journal of Chromatography A*, 720(1-2):295–321, 1996.

[159] E Staudacher, F Altmann, I B Wilson, and L Mrz. Fucose in n-glycans: from plant to man. *Biochimica Et Biophysica Acta*, 1473(1):216–236, December 1999. PMID: 10580141.

[160] G Stevenson, K Andrianopoulos, M Hobbs, and PR Reeves. Organization of the escherichia coli k-12 gene cluster responsible for production of the extracellular polysaccharide colanic acid. *J. Bacteriol.*, 178(16):4885–4893, August 1996.

[161] Stoyanka Stoitsova, Radka Ivanova, and Ivanka Dimova. Lectin-binding epitopes at the surface of escherichia coli k-12: examination by electron microscopy, with special reference to the presence of a colanic acid-like polymer. *Journal of Basic Microbiology*, 44(4):296–304, 2004.

[162] Christine M Szymanski, Ruijin Yao, Cheryl P Ewing, Trevor J Trust, and Patricia Guerry. Evidence for a system of general protein glycosylation in campylobacter jejuni. *Molecular Microbiology*, 32(5):1022–1030, June 1999.

[163] Lisa Tang, Adam M Persky, Gnther Hochhaus, and Bernd Meibohm. Pharmacokinetic aspects of biotechnology products. *Journal of Pharmaceutical Sciences*, 93(9):2184–2204, September 2004.

[164] Sheng-Ce Tao, Yu Li, Jiangbing Zhou, Jiang Qian, Ronald L Schnaar, Ying Zhang, Irwin J Goldstein, Heng Zhu, and Jonathan P Schneck. Lectin microarrays identify cell-specific and functionally significant cell surface glycan markers. *Glycobiology*, 18(10):761 –769, October 2008.

[165] Anthony L. Tarentino, Thomas H. Plummer Jr., and Gerald W. Hart William J. Lennarz. Enzymatic deglycosylation of asparagine-linked glycans: Purification, properties, and specificity of oligosaccharide-cleaving enzymes from flavobacterium meningosepticum. In *Guide to techniques in glycobiology*, volume Volume 230, pages 44–57. Academic Press, 1994.

[166] Christopher J. Thibodeaux, Charles E. Melan|[ccedil]|on, and Hung-wen Liu. Unusual sugar biosynthesis and natural product glycodiversification. *Nature*, 446(7139):1008–1016, April 2007.

[167] Juan D Valderrama-Rincon, Adam C Fisher, Judith H Merritt, Yao-Yun Fan, Craig A Reading, Krishan Chhiba, Christian Heiss, Parastoo Azadi, Markus Aebi, and Matthew P DeLisa. An engineered eukaryotic protein glycosylation pathway in escherichia coli. *Nature Chemical Biology*, March 2012.

[168] M. A. Valvano. Common themes in glycoconjugate assembly using the biogenesis of o-antigen lipopolysaccharide as a model system. *Biochemistry (Moscow)*, 76:729–735, July 2011.

[169] D. M.F van Aalten and N. C.J Strynadka. Putting glycobiology on a structural footing editorial overview. *Current Opinion in Structural Biology*, 18:525526, 2008.

[170] Scott M. Van Patten, Heather Hughes, Michael R. Huff, Peter A. Piepenhagen, James Waire, Huawei Qiu, Chandrashekar Ganesa, David Reczek, Paul V. Ward, Joseph P. Kutzko, and Tim Edmunds. Effect of mannose chain length on targeting of glucocerebrosidase for enzyme replacement therapy of gaucher disease. *Glycobiology*, 17(5):467 –478, May 2007.

[171] Ajit Varki, Richard D. Cummings, Jeffrey D. Esko, Hudson H. Freeze, Gerald W. Hart, and Marilynn E. Etzler. *Essentials of Glycobiology, Second Edition*. Cold Spring Harbor Laboratory Press, 2nd edition, October 2008.

[172] Michael Wacker, Mario F. Feldman, Nico Callewaert, Michael Kowarik, Bradley R. Clarke, Nicola L. Pohl, Marcela Hernandez, Enrique D. Vines, Miguel A. Valvano, Chris Whitfield, and Markus Aebi. Substrate specificity of bacterial oligosaccharyltransferase suggests a common transfer mechanism for the bacterial and eukaryotic systems. 103(18):7088–7093, May 2006.

[173] Michael Wacker, Dennis Linton, Paul G. Hitchen, Mihai Nita-Lazar, Stuart M. Haslam, Simon J. North, Maria Panico, Howard R. Morris, Anne Dell, Brendan W. Wren, and Markus Aebi. N-Linked glycosylation in campylobacter jejuni and its functional transfer into e. coli. *Science*, 298(5599):1790–1793, November 2002.

[174] S. Wagner. *From Biogenesis to Overexpression of Membrane Proteins in Escherichia coli*. Department of Biochemistry and Biophysics, Stockholm university, 2008.

[175] Samuel Wagner, Louise Baars, A. Jimmy Ytterberg, Anja Klussmeier, Claudia S. Wagner, Olof Nord, Per-Ake Nygren, Klaas J. van Wijk, and Jan-Willem de Gier. Consequences of membrane protein overexpression in escherichia coli. *Mol Cell Proteomics*, 6(9):1527–1550, September 2007.

[176] Samuel Wagner, Mirjam M Klepsch, Susan Schlegel, Ansgar Appel, Roger Draheim, Michael Tarry, Martin Hgbom, Klaas J van Wijk, Dirk J Slotboom, Jan O Persson, and Jan-Willem de Gier. Tuning escherichia coli for membrane protein overexpression. *Proceedings of the National Academy of Sciences of the United States of America*, 105(38):14371–6, September 2008. PMID: 18796603.

[177] Xu Wang, Thomas Weldeghiorghis, Guofeng Zhang, Barbara Imperiali, and James H. Prestegard. Solution structure of alg13: The sugar donor subunit of a yeast N-Acetylglucosamine transferase. *Structure*, 16(6):965–975, June 2008.

[178] Eranthie Weerapana and Barbara Imperiali. Asparagine-linked protein glycosylation: from eukaryotic to prokaryotic systems. *Glycobiology*, 16(6):91R–101, June 2006.

[179] Thomas Weide, Lutz Herrmann, Ulrike Bockau, Nadine Niebur, Ingo Aldag, Wouter Laroy, Roland Contreras, Arno Tiedtke, and Marcus WW Hartmann. Secretion of functional human enzymes by tetrahymena thermophila. *BMC Biotechnology*, 6:19, March 2006. PMID: 16542419 PMCID: 1431531.

[180] S. Weikert, D. Papac, J. Briggs, D. Cowfer, S. Tom, M. Gawlitzek, J. Lofgren, S. Mehta, V. Chisholm, N. Modi, S. Eppler, K. Carroll, S. Chamow, D. Peers, P. Berman, and L. Krummen. Engineering chinese hamster ovary cells to maximize sialic acid content of recombinant glycoproteins. *Nat Biotech*, 17(11):1116–1121, November 1999.

[181] Gavin J. Williams, Randal D. Goff, Changsheng Zhang, and Jon S. Thorson. Optimizing glycosyltransferase specificity via Hot spot saturation mutagenesis presents a catalyst for novobiocin glycorandomization. *Chemistry & Biology*, 15(4):393–401, April 2008.

[182] Gavin J Williams and Jon S Thorson. A high-throughput fluorescence-based glycosyltransferase screen and its application in directed evolution. *Nat. Protocols*, 3(3):357–362, March 2008.

[183] Gavin J Williams, Changsheng Zhang, and Jon S Thorson. Expanding the promiscuity of a natural-product glycosyltransferase by directed evolution. *Nature Chemical Biology*, 3(10):657–662, September 2007.

[184] William C. Wimley and Stephen H. White. Designing transmembrane - Helices that insert spontaneously. *Biochemistry*, 39(15):4432–4442, April 2000.

[185] Sima Yaron, Glynis L. Kolling, Lee Simon, and Karl R. Matthews. Vesicle-Mediated transfer of virulence genes from escherichia coli O157:H7 to other enteric bacteria. *Appl. Environ. Microbiol.*, 66(10):4414–4420, October 2000.

[186] Elif Yavuz, Carola Maffioli, Karin Ilg, Markus Aebi, and Bernard Priem. Glycomimicry: display of fucosylation on the lipo-oligosaccharide of recombinant escherichia coli k12. *Glycoconjugate Journal*, 28(1):39–47, January 2011. PMID: 21286806.

[187] N Martin Young, Jean-Robert Brisson, John Kelly, David C Watson, Luc Tessier, Patricia H Lanthier, Harold C Jarrell, Nicolas Cadotte, Frank St Michael, Erika Aberg, and Christine M Szymanski. Structure of the n-linked glycan present on multiple glycoproteins in the gram-negative bacterium, campylobacter jejuni. *The Journal of Biological Chemistry*, 277(45):42530–42539, November 2002. PMID: 12186869.

[188] Shin-Yi Yu, Sz-Wei Wu, and Kay-Hooi Khoo. Distinctive characteristics of MALDI-Q/TOF and TOF/TOF tandem mass spectrometry for sequencing of permethylated complex type n-glycans. *Glycoconjugate Journal*, 23(5-6):355–369, July 2006.

[189] H. Zappe, M.E. Snell, and M.J. Bossard. PEGylation of cyanovirin-N, an entry inhibitor of HIV. *Advanced Drug Delivery Reviews*, 60(1):79–87, January 2008.

[190] T. Zheng, H. Jiang, M. Gros, D. Soriano del Amo, S. Sundaram, G. Lauvau, F. Marlow, Y. Liu, P. Stanley, and P. Wu. Tracking N-Acetyllactosamine on Cell-Surface glycans in vivo. *Angewandte Chemie International Edition*, 50(18):41134118, 2011.